nature

www.nature.com/nature

Vol 449 | Issue no. 7159 | 13 September 2007

Peer review reviewed

The US research community is responding vigorously to calls to help change the system of grant assessment at the National Institutes of Health. A radical transformation is urgently needed.

The peer-review system used by the $29-billion National Institutes of Health (NIH) is more than half-a-century old, and is showing its age. It has become stretched by the breadth of today's science, in which inter- and multidisciplinary grant applications are common, and by the sheer volume of submissions in an era in which one-grant labs have gone the way of the dinosaur.

Twenty years ago, some 1,800 reviewers judged grant applications for the NIH's Center for Scientific Review, which oversees the lion's share of the agency's peer review; today, that number is more than 18,000. Increasingly, ad hoc and junior reviewers have been called into service — the former to provide expertise on complex multidisciplinary grants, the latter because of volume and because senior scientists feel that they have already paid their dues with earlier service.

Complicating this unwieldy situation, the current NIH funding freeze has made funding committees conservative to the point that an application must be almost perfect to be funded on its first submission. A searching assessment of how the system can be reshaped and improved is essential. The Center for Scientific Review has already sought feedback through a series of field-specific community sessions and has tested some changes, such as shorter grant-review cycles and a more electronic grant-evaluation process.

Separately, NIH director Elias Zerhouni in June launched a bid to restructure peer review at the NIH to reflect foreseeable needs. An internal panel of senior NIH officials and an illustrious working group of non-NIH scientists jointly face a deceptively simple challenge: to ensure that the agency funds the best science by the best scientists with the lightest administrative burden.

Ideas were solicited this summer at a packed meeting of scientific-society leaders in Washington DC. More researchers' opinions are being gathered at meetings in Chicago, New York and San Francisco this month and next. Electronic comments were invited over a two-month period that ended last Friday. Judging by the 2,000 opinions submitted, the extramural community has plenty to say on the matter.

The two groups aim to have concrete recommendations by early this winter, with the goal of launching pilot projects as soon as next spring. They have asked for 'creative' and even 'radical' ideas, intending to act not on the most popular suggestions but on the best ones.

Still, good ideas have emerged in the 'popular' category: there are strong arguments to be made for shortened grant applications and for regular 'bridge' funding to see investigators through gaps between grants. It is also important to ensure that senior, accomplished scientists serve on study sections. There is simply no replacement for the brains, experience, insight and judgement that they bring to bear on applications.

"The NIH needs to fund the best science by the best scientists with the lightest administrative burden."

To this end, NIH grantees should be required to serve on study sections if the agency asks for their help — with due provision to ensure that they are not overburdened, and perhaps also a reward in the form of increased funds for their own grants. This would ensure that the best scientists are recruited onto study sections, and that senior scientists are brought back into the system.

A proposal put forward by the Association of American Medical Colleges, and probably others, would allow individual scientists to have only one application of a given kind in the system at any one time. Multiple grants could still be held by one scientist, but he or she could have only one application per mechanism under review. This would compel self-selection of the best proposals by scientists upstream of the review process. To be workable, this would necessitate a funding cycle that lasts at most six months rather than the current ten. But that compression is highly desirable in any case and has already been accomplished in pilot trials.

Such an approach can only help the most creative scientists by stemming the current deluge of applications. It's a radical idea but, for that reason at least, an excellent one. ■

Meeting obligations

Climate change should take ever-increasing priority in the Asia-Pacific region.

Gatherings of world leaders are never easy events, and last week's Asia-Pacific Economic Cooperation (APEC) forum in Sydney, Australia, was no exception. The United States and South Korea, for example, shared some awkward moments over whether the Korean War should officially be declared over; and environmental activists complained that not enough was done to advance one of the meeting's key issues: climate change.

Yet the very fact that climate change was on the APEC agenda was a start. It was put there by one of the environmentalists' greatest foes, Australian prime minister John Howard — a man who has consistently opposed the notion of mandatory emissions cuts. Unsurprisingly, the statement signed by the 21 APEC leaders was vague, calling for just two specific actions: an additional 20 million hectares of forest in the region by 2020, and a 25% reduction in energy intensity — the amount of greenhouse gases released per dollar of gross domestic product — by 2030. And there are no penalties set out for not meeting these 'aspirational' goals.

It is encouraging that the APEC leaders have issued a climate consensus, however weak. Such discussions, after all, emphasize the increasing importance that the Asia-Pacific region plays in the

climate-change arena. Too often the United States and Europe are portrayed as the main players on climate issues, while Asian countries feature mainly when others excuse their alleged inaction by pointing fingers at the booming economies of China and India, who under the Kyoto Protocol on climate change are not bound to reduce their emissions. But China is moving ahead on its own — President Hu Jintao has regularly spoken about the importance of climate change as a global issue, and last week his country announced plans to get 15% of its energy from renewable sources by 2020.

Political changes in some of the countries holding out on climate change may help facilitate Asian action. Howard is expected to call elections for this winter, and he is running far behind his opposition in the polls. George W. Bush will be out as of January 2009, and nearly all of the leading presidential candidates could provide the US leadership on climate change that has been so sorely lacking.

So what next? Yet more meetings. Earlier this week a number of the Asian players, including Australia, China, Indonesia and India, joined the 'Gleneagles dialogue' in Berlin, in which energy and environment ministers discuss clean-energy goals. This is but a minor step on the path to a real emissions policy; another such sidestep will come at the end of this month, when Bush launches discussions in Washington DC on what to do about climate-change targets when the Kyoto agreement expires in 2012. As the United States has not ratified Kyoto, this is likely to be something of a distraction.

Stakeholders should instead focus their efforts on the talks in early December in Bali, Indonesia, which will include all the parties to Kyoto. This meeting, run by the United Nations Framework Convention on Climate Change, embodies the de facto international framework for discussing climate change, and as such is the outlet best suited for constructing emissions commitments.

International negotiators must work together towards a clear and consistent discussion at all these meetings. Representatives from the Asian bloc should continue to keep climate change as a high priority, and make more aggressive moves towards implementing real targets for emissions cuts at the Bali meeting. Asia has both the economic clout and the incentive to be a world leader in climate change. ■

Turkey's transformation

A European vision and a commitment to openness will foster good science.

Turkish scientists have never had it so good, thanks to their country's efforts to align its laws and policies to those required for membership of the European Union (EU). In a bid to create a science and higher-education landscape that matches the EU norm, Turkey has more than doubled its research spending in the past five years, and is half way to its goal of spending 2% of its gross domestic product on research by 2010. It has refined its peer-review procedures for research grants to improve fairness and transparency, and is actively promoting research in industry. The country's best scientists say that for the first time it is now possible to get grants of a decent size — even up to hundreds of thousands of dollars — for a strong basic-research project.

To be able to spend the new money as wisely as possible, Turkey needs to expand, and rejuvenate, its relatively small community of scientists. Plans are in motion, thanks again to the country's westward focus. Nineteen new universities will be founded in the next few years. Special grants to allow young scientists to set up independent research labs in universities have been established. And with so much more money available for research, Turkish scientists are now starting to come home from abroad.

To encourage individual scientists to become more active, Tubitak, the main research agency — and sometimes the universities themselves — top up the personal salaries of grant-winners and offer financial incentives for publication in international journals. This has helped push Turkey up from 27 to 19 in the world rank of science publication rates since 1997.

But impact, as measured by citations per paper, has increased only slightly in that time. And Turkey's commitment wavered last year after its scientists won few grants from the sixth EU Framework programme (2002–06) — the first to which it contributed funds. It was persuaded only with difficulty to join the seventh programme. Finally accepting that its scientists would only improve in the short term through continued close contact and competition with their EU colleagues, the Turkish government signed on the dotted line in June.

All seems to be set fair for scientific growth — provided Turkish politics remain stable. This seems likely, despite alarm bells being sounded by many inside and outside the country following the recent re-election of a mildly Islamic government, headed by a new religiously minded president. The old-guard academic élite, in particular, foresee dire consequences arising from the almost certain relaxation of rules that ban headscarves in government-funded institutions, including universities. What they fear most is an 'Islamization' of politics, and the discarding of the secular constitution written by Mustafa Kemal Atatürk when he founded the Republic of Turkey in 1923. The headscarf is a powerful symbol of the tensions between ardent secularists and the religious — tensions that are evident in all areas of public life.

But more scientists are coming to accept that the right to wear a headscarf in an educational establishment may not, after all, be the thin end of an extremist wedge. Secularism's deep roots won't be overturned so easily in a country where the majority of the electorate, who happen to be religious, are becoming materially wealthier under Western aspirations. There is no room for complacency, however. Frictions closer to the country's eastern border with Iran — where a university rector recently fell foul of religious groups and ended up in jail — are less easy to control. Nationalism is also a threat to stability; an insidious law criminalizing 'insulting Turkishness', which has been used on occasion to silence public opposition, needs to be repealed.

Science was the first focus of negotiations in Turkey's bid for EU membership — and had been so well prepared that the chapter could be closed in only nine months. The nation's European ambitions are also likely to provide an incentive for repeal of the nationalist law. Whether or not Turkey will become the EU's first Muslim member state is hard to predict, but the benefits of that ambition for science and more are clear. ■

RESEARCH HIGHLIGHTS

I. KATO/REUTERS

Runner gene

Nature Genet. doi:10.1038/ng2122 (2007)

A mutation commonly found in endurance athletes may have been favoured by evolution because it aids efficient muscle function, say researchers in Australia. More than a billion humans worldwide are predicted to have the mutation, causing them to lack a 'fast' muscle-fibre protein known as α-actinin-3. Absence of this protein seems to boost stamina, as metabolic resources are diverted onto a slower but more efficient metabolic pathway.

Researchers led by Kathryn North of the Children's Hospital at Westmead in Sydney found that mice lacking α-actinin-3 ran on average 33% farther on a treadmill than normal mice before reaching exhaustion. They also show that the human version of the mutation is surrounded by well-conserved DNA sequence, suggesting that it has been favoured by natural selection.

NEUROSCIENCE

A glimpse of the impossible

Nature Neurosci. doi:10.1038/nn1951 (2007)

Researchers have proposed an explanation for why our brain can detect a visual phenomenon that doesn't naturally occur.

The brain processes images from each eye to establish a single picture. It registers differences between the relative positions of objects in each eye to glean depth information. But the visual cortex also has neurons that register an improbable effect known as 'phase disparity' — differences in the patterns of light and dark.

Jenny Read of Newcastle University, UK, and Bruce Cumming of the National Eye Institute in Bethesda, Maryland, hypothesized that the ability to detect phase disparity allows the brain to recognize when it has incorrectly aligned the eyes' two images. They applied a computer model of how the brain may do this to a stereogram of the Pentagon, headquarters of the US Department of Defense. The model yielded reasonably accurate images (pictured right), outperforming models that lack checks on phase disparity.

PHYSICS

Iced neutrons

Phys. Rev. Lett. **99,** 104801 (2007)

Good news for physicists who like their neutrons chilled: Oliver Zimmer of the Technical University of Munich in Germany and his co-workers have demonstrated the viability of one long-standing proposal for making 'ultra-cold neutrons'.

Such neutrons, which move at no more than human running speed, can probe fundamental aspects of physics, for example, the decay lifetime of the neutron itself. But the best nuclear-reactor sources offer only a few dozen ultracold neutrons in each thimbleful of space — too few for easy study.

Zimmer and his colleagues cooled neutrons from a research reactor source by passing them through superfluid helium. Such cooling has been achieved before, but the researchers have now also shown how to accumulate the ultracold neutrons before extracting them to attain higher neutron densities.

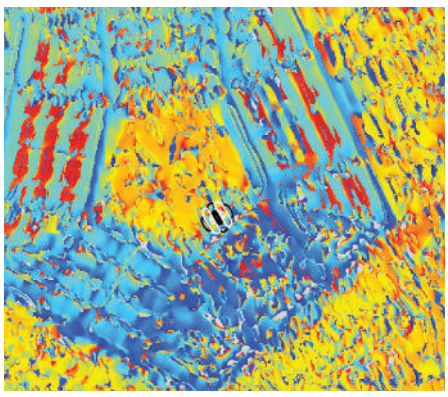
EARTH SCIENCE

No oxygen required

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0704912104 (2007)

It is widely accepted that there was not a persistent, significant amount of oxygen in Earth's atmosphere more than 2.45 billion years ago. But in 1999, hydrocarbon molecules called 2-methylhopanes, thought to be biomarkers distinctive of oxygen-producing cyanobacteria, were found in sediments that are 2.7 billion years old. This has led to much discussion of how an oxygen-producing biosphere could persist for hundreds of millions of years before any of its oxygen accumulated in the atmosphere.

Sky Rashby of the California Institute of Technology in Pasadena and his colleagues argue that 2-methylhopanes may not be de facto evidence for oxygenic photosynthesis. They found that the purple non-sulphur bacterium *Rhodopseudomonas palustris* produces 2-methylhopanes. Although this bacterium, like cyanobacteria, is photosynthetic, it does not produce oxygen, and it needs no oxygen in order to make the biomarkers.



BIOCHEMISTRY

Radical weapons

Cell **130,** 797–810 (2007)

Antibiotics that target different cellular pathways have been found to have a common mode of killing. The lethal weapons, identified by James Collins and his colleagues at Boston University in Massachusetts, are hydroxyl free radicals — highly reactive molecules that damage DNA and proteins.

Using a hydroxyl-sensitive fluorescent dye, the researchers showed that antibiotics that target bacterial cell-wall, protein or DNA synthesis pathways trigger the production of hydroxyl radicals. They also found that blocking production of these radicals decreases the potency of antibiotics, whereas

blocking the DNA-repair pathway that bacteria use to fight free-radical damage increases cell death. The team says that the findings point to routes to enhancing the potency of antibiotics.

CHEMISTRY

Supercentre

Org. Lett. doi:10.1021/ol701911u (2007)

Making 'stereogenic centres' in molecules — regions where the arrangements of atoms can take left- and right-handed forms — is one of the toughest problems for synthetic chemists. It is usually hard enough to make a single such centre with the desired handedness, but Jingqiang Wei and Jared Shaw of the Broad Institute of Harvard and Massachusetts Institute of Technology in Cambridge have achieved the feat of creating two or three of them at once. Even more impressively, they do it by combining four separate molecular building-blocks in a single step. The result is a compound called a γ-lactam, which has a ring of four carbon atoms and one nitrogen and is a useful potential 'backbone' molecule for making new drugs.

CELL BIOLOGY

Fresh packaging

Nature Cell Biol. doi:10.1038/ncb1636 (2007)

New observations may settle controversy about how the membrane that wraps a cell's nucleus regenerates after cell division.

A replicating cell must break apart this membrane, known as the nuclear envelope, to split its chromosomes between its daughter cells. Daniel Anderson and Martin Hetzer of the Salk Institute for Biological Studies in La Jolla, California, show that the membranes in the daughter cells derive not from fragments of the old membrane, as one theory held, but from an internal network of tubular membranes called the endoplasmic reticulum (ER).

Fluorescent labelling of the ER allowed the researchers to watch the tubules expand and flatten into membrane sheets around the nucleus. They also showed that antibodies against a protein required for ER tubule formation inhibited development of the nuclear envelope.

ASTRONOMY

Double trouble

Astrophys. J. **666,** L89–L92 (2007)

Observations of stars shooting away from the Milky Way's core could reveal whether there is more than one black hole lurking at the Galaxy's centre, scientists say.

Most researchers agree that at least one massive black hole sits at the centre of the Galaxy (pictured, below), but if the Milky Way assembled through the merger of smaller galaxies, it may have two or more central black holes. Calculations by Youjun Lu and his colleagues at the University of California, Santa Cruz show that, if a binary black hole system survives at the Milky Way's centre, it could be identified through its distinctive effect on binary stars, ejecting one pair every 200,000 years. Astronomers have already seen lone 'hypervelocity' stars thought to have been kicked out of the Galactic core; now the team suggests a search for binaries.

NASA/CXC/MIT/F. K. BAGANOFF *ET AL.*

ECOLOGY

Breaking the cycle

PLoS Biol. **5,** e239 (2007)

Lab experiments have unearthed an exception to a classic tenet of ecology. The experiments show that the size of a predator population doesn't always cycle in tandem with the size of the population of its prey — a finding that may force ecologists to rethink how they search for species interactions.

Nelson Hairston at Cornell University in Ithaca, New York, and his colleagues first noticed the odd behaviour among rotifers and their algal prey — the number of algae would sometimes stay steady while rotifer numbers waxed and waned. This happened when the single species of alga was represented by several genotypes, some better at resisting rotifers than others. The researchers also observed similar dynamics between a bacterium and a virus.

A mathematical model suggests that species interactions become 'hidden' when the prey can rapidly evolve defence mechanisms against the predator and the evolutionary cost of doing so is low.

JOURNAL CLUB

Julian Davies
University of British Columbia, Vancouver, Canada

A microbiologist wonders where diversity comes from.

Recent estimates indicate that the total number of bacteria in the biosphere approaches or exceeds 10^{31}. A major goal of microbiology is to understand what creates their diversity and how it is maintained.

Having trained as an organic chemist, I came to appreciate microbial diversity through the extravagance of small molecules that microbes produce. This reflects a diversity in microbial metabolism, which one might expect to have evolved as a result of the (organic) richness of the organisms' environments. But a couple of recent publications present findings that do not sit easily with this view.

Our first inkling of the huge diversity of the microbial world came from the use of ribosomal-RNA typing in the late 1980s. In the 1990s, this morphed into the expanding field of metagenomics, which is now providing catalogues of microbial communities from diverse terrestrial and marine environments.

One comparison of such catalogues showed that the seemingly bare and boring Arctic tundra exceeds fertile forest soils in phylogenetic content (J. D. Neufeld and W. W. Mohn *Appl. Environ. Microbiol.* **71,** 5710–5718; 2005). A more recent study compared information from more than 100 different environments, finding that the microbial content of soils is generally less diverse than that of sediments and hypersaline environments (C. A. Lozupone and R. Knight *Proc. Natl Acad. Sci. USA* **104,** 11436–11440; 2007).

I am looking forward to seeing what happens when the Human Microbiome Project gets under way. What variety of microbes is there to find living within us? What are they all doing? In what way will the population depend on diet? Given that we don't yet seem to understand the relationship between diversity and ecology, I am making no predictions.

Discuss these papers at **http://blogs.nature.com/nature/journalclub**

NEWS

There is a correction associated with this article (see overleaf).

Borysiewicz to head UK medical council

The next head of the UK Medical Research Council (MRC) has been chosen, potentially bringing to a close a period of significant unrest within the organization. *Nature* has learned that the government's biomedical research organization will appoint Leszek Borysiewicz as successor to its current chief executive Colin Blakemore when he steps down at the end of this month.

Borysiewicz, who is currently deputy rector at Imperial College London, has a clinical-sciences background and is currently co-chair of the MRC's advisory group on stem-cell research. He will take charge of the MRC's £460-million (US$933-million) annual research budget — perhaps more, if a government decision to increase the MRC's spending power is announced as expected later this year. "There will be a very substantial increase in the MRC's budget," says Blakemore. "I think the MRC will retain and expand its work in fundamental research, but will also expand elsewhere."

Borysiewicz will inherit an organization whose scientists are beset with doubts about their future. Last year's government-commissioned review of the MRC's goals by venture capitalist and former Wellcome Trust governor David Cooksey called for the council to pursue an agenda of 'translational research' — biomedical research more strongly focused on health benefits and the economic bottom line. The review voiced economists' and executives' fears that, despite the MRC's impressive track record of medical discoveries such as cancer drugs and monoclonal antibodies, it has failed in the past to maximize the clinical impact and reap the cash rewards its innovations deserve. But the new agenda has left MRC-funded scientists worried that basic research will be left out in the cold.

"I'm a lot more confident than I was six months ago that the MRC will not change its commitment to fundamental research across the whole spectrum, but in addition will be given the ability through funding to build more efficiently and carry forward things towards application," Blakemore told *Nature*.

Critics of the translational research agenda are anxious that the new appointee should not stifle basic research. "We need someone committed to making sure that 'blue sky' research funding is maintained," says Hilary Leevers, acting director of the Campaign for Science and Engineering in the UK.

IMPERIAL COLLEGE

Leszek Borysiewicz has a strong background in biomedical research.

"We need to make sure science is balanced." With a strong basic research background that has also yielded clinical benefits such as vaccines, Borysiewicz looks well placed to deliver on this balancing act.

The process of selecting a new chief executive has also suffered controversy, with doubts about the suitability of MRC chairman John Chisholm — appointed last year to lead the search for a new chief executive — to select the right candidate. In July, the Commons Science and Technology Committee said that it had "serious reservations as to whether Sir John is the right person to guide the MRC executive through the coming period of change". Chisholm previously presided over the spin-out of the government's defence research agency to form a profit-making company called QinetiQ. At the MRC, he was appointed as a non-executive chairman with no involvement in the council's decision on where to direct funds, but rumours have persisted that more than one well-qualified candidate has been discouraged from applying for the role of chief executive owing to fears of interference in such decisions. Blakemore says he was not aware of candidates declining to apply, adding that "I am very confident that the next chief executive will be the leader of the MRC." Although Blakemore claims that he has no knowledge of who will replace him, he says: "I'm confident we will have a new chief executive in the post at the time of my departure." He adds that selecting a chief executive from a clinical background would potentially be a good strategic move.

> "The appointment of someone of high standing and professional reputation is crucial."
> — Martin Rees

"The most important thing is that the MRC should, and I think will, maintain its quality of judgement in supporting the very best biomedical research in the UK," Blakemore says.

Most observers of British science agree that the translational research agenda is a necessary and pragmatic new direction, providing that basic science does not suffer unduly. This was a widespread fear when Blakemore announced his intention to leave earlier this year. "We hope the agenda can be implemented without losing research quality," says Royal Society president Martin Rees. "The appointment of someone of high standing and professional reputation is crucial."

Borysiewicz, described by Rees as a "distinguished figure", has a research background focused on viruses and immunology. In 2001 he received a knighthood for his work on developing a range of vaccines, including the vaccine against human papilloma virus aimed at preventing cervical cancer. A popular figure among students and researchers at Imperial College, Borysiewicz is responsible for the college's overall scientific and academic direction. He has embraced applied research, particularly in establishing the Schistosomiasis Control Initiative, which was funded by the Bill & Melinda Gates Foundation.

Another influential British science policy job also looks to have been given to a candidate from Imperial College. John Beddington, a biological economist and political adviser on fisheries, is to be asked to become the British government's new chief science adviser, replacing David King when he finishes his eight-year tenure at the end of this year. The Department of Innovation, Universities and Skills is expected to make a formal announcement shortly.

With a background in environmental and fisheries research, Beddington is well-versed in the issues that look set to dominate government science policy during the next few years. "We at the Royal Society feel he's an excellent choice," says Rees. ■

Michael Hopkin

Correction

In the News story 'Borysiewicz to head UK medical council' (*Nature* **449,** 121; 2007), we misquoted Colin Blakemore, chief executive of the Medical Research Council (MRC), in a way that suggested he knew that there would be a substantial increase in the MRC's budget. Professor Blakemore said that he hoped there would be a substantial increase in the MRC's budget, and points out that he is not in a position to declare a definite funding increase. We apologize to Professor Blakemore.

Russian scientists see red over clampdown

A young Russian biologist who was taking samples to a collaborative institute in France has been accused of attempting to smuggle bioweapons by Russia's federal security service, the FSB. He has been interrogated repeatedly by FSB agents and prevented from leaving the country. His job also now looks uncertain. But experts say that the accusations are absurd.

Oleg Mediannikov's Kafkaesque nightmare began on 12 December 2006 at Moscow's Sheremetyevo airport, as he was about to board a plane to Marseilles. Customs officials confiscated 20 phials containing non-pathogenic strains of a typhus vaccine approved by the Russian health ministry for export to France, along with Mediannikov's computer and USB memory sticks. Mediannikov initially thought there was a minor problem with the paperwork. But more than eight months on, the interrogations continue.

Mediannikov, who works at the Gamelaya Institute of Epidemiology and Microbiology in Moscow, studies *Rickettsia prowazekii*, the bacterium spread by lice that causes epidemic typhus. The institute's laboratory of rickettsial ecology, headed by Irina Tarasevich, has a long-established collaboration with the Rickettsial Unit of the University of the Mediterranean in Marseilles, led by Didier Raoult. The two institutes are World Health Organization Collaborating Centres for Rickettsial Reference and Research.

"If things get worse, we will demand that the FSB interrogates the deputy health minister, who approved the export of the material."
— Oleg Mediannikov

Raoult planned to compare the protein spectrum of two strains of *R. prowazekii* — Madrid E and EVir — produced more than 20 years ago in chicken embryos in Russia and since held at the Gamelaya Institute, with similar strains produced more recently in France from mammalian cell cultures. Both strains are not considered to be virulent and are used in vaccinations against typhus. The work is part of a larger research project on the pathogenesis of *R. prowazekii*, led by Raoult and funded by France's basic-research agency, the CNRS.

Mediannikov was allowed to continue his trip to Marseilles without the samples. On his return to Moscow in January, he was told that the confiscated material had been sent to a secret laboratory — code-named the 47th military research institute — for an 'expert assessment'. Three weeks later he was told that an additional assessment — the first allegedly concluded the materials were benign — was necessary before the materials could be returned. This second assessment is still pending.

But the situation is causing other problems for Mediannikov. On 13 February, he intended to go on a tourist trip to Cameroon, only to learn at Moscow's Domodedovo airport that there was an official order preventing him from leaving the country. When he demanded an explanation, a customs official said the order "must not be discussed". His passport was confiscated and returned two months later by regular post.

All his efforts to clarify the situation have proved fruitless. In early June, customs informed him that the FSB — successor to the Soviet KGB — insisted on initiating criminal proceedings. To avert prosecution, he gave them the valid export permission signed by the deputy health minister. In addition, he presented letters from Tarasevich and Raoult attesting to the harmlessness of the strains and their sole use for scientific purposes.

Nonetheless, criminal proceedings were initiated on 26 June — and the accusations are severe. The indictment, of which *Nature* has obtained a copy, cites Article 188/2 of the Criminal Code of the Russian Federation on smuggling materials that might be used for preparing weapons of mass destruction. People guilty of illicit trafficking of weapons-delivery systems can be sentenced to up to seven years in prison.

Raoult says that he is stunned. "Something like this has never happened in 20 years of our centres' collaboration," he says. "Oleg spent two years working in my lab. He is a very good, dynamic and responsible scientist." His work has been instrumental in helping fight typhus in Russia, he adds. Typhus bacteria are not considered potential bioterrorism agents by other governments. "It is a terrible disease, but the agent is so difficult to grow that it doesn't make any sense to use it for bioweapons," says Raoult.

This week, Mediannikov told *Nature* that the deputy director of his institute had been approached by the FSB and that Mediannikov has now been told to resign or face the sack.

On the up

Mediannikov's case illustrates a worrying resurgence in Russian scientists being accused of wrong-doing. In 2000, for example, thermophysicist Valentin Danilov of Krasnoyarsk State Technical University was arrested for allegedly passing classified information to China. He was acquitted in 2003, but taken into custody again after the Russian Supreme Court overturned the acquittal in 2004. And in 2004, Igor Sutyagin, a social scientist formerly with the US and Canada Institute in Moscow, was sentenced to 15 years in a labour camp for allegedly passing classified data on nuclear submarines and missile-warning systems to a British company.

The FSB also suspected chemist Oleg Korobeinichev, head of the laboratory of chemical kinetics and combustion in Novosibirsk, of having divulged state secrets to the United States. But the charges were dropped in June 2006, and Korobeinichev received a public apology from local legal authorities. On 27 August, the FSB finally withdrew the charges against physicists Oleg and Igor Minin, who had been accused of revealing state secrets.

"There have been worse times in this coun-

M. MARMUR/AFP/GETTY

NSTITUT PASTEUR/PHOTOTAKE INC./PHOTOLIBRARY

Russia's FSB (above) has detained several scientists for 'smuggling' scientific samples such as *Rickettsia prowazekii* (right) out of the country.

try," says a Russian expert on non-proliferation on condition of anonymity. "But Vladimir Putin has untied the hands of the FSB, and we do see a trend here towards strengthening state control over all spheres of life, including science."

In May, the FSB warned in a secret report to President Putin that biological samples taken from Russians could be used abroad to produce 'genetic weapons'. Consequently, the export of human specimens was temporarily banned. The order was reversed two weeks later after an outcry in the media and the scientific community.

"Publicity does help in such cases," says Konstantin Severinov, a biochemist who has a joint affiliation at the Institute of Molecular Genetics in Moscow and at Rutgers University in New Jersey. Severinov was himself 'interviewed' by an FSB official in June. "I told the guy straight away that the whole genetic-weapon craze is nothing but lunacy and paranoia," he says.

Over the past couple of months, Mediannikov has been summoned six times to the FSB interrogation department in Moscow. Interviews — about his biography, scientific advisers, collaborators, research, and so on — lasted for up to four hours, but took place in a "quite pleasant" atmosphere, he says. Mediannikov is now waiting for the result of the second expert assessment. "If there's anything in it that might back the charge we will insist on a third, independent assessment," he says. "If things get worse, we will also demand that the FSB interrogates the deputy health minister, who approved the export of the material." He points out that scientists from the Gamelaya Institute have previously taken similar samples to France without any problems.

One customs officer, Mediannikov says, hinted to him that customs were "ordered" by the FSB to take action against him. And Severinov says that Mediannikov might have been shopped to the FSB by an over-zealous member of his institute's 'first department'. These notorious 'security' departments are obligatory at Russian research institutes — a relic of Soviet times — and they maintain close connections with the FSB.

Mediannikov's situation is serious, as is always the case when FSB investigators are involved, say legal experts. But if convicted only of 'ordinary' smuggling, he may yet get away with a modest penalty fee, they say. A date for the trial has not been set. He is not in custody, but experts doubt that he would be allowed to leave Russia as long as the investigation continues. ■

Quirin Schiermeier

Mystery ox finds its identity

The kouprey, an enigmatic Asian ox believed to be a hybrid — and so, unworthy of conservation efforts — is in fact a distinct species related to the banteng (a wild ox)[1]. The conclusion contradicts earlier findings[2] that the horned beast is a cross between the banteng and domesticated zebu cattle.

First identified in 1937 and last spotted in the 1980s, the kouprey (*Bos sauveli*) has become a symbol for conservation in southeast Asia. Some experts think that it is already extinct.

Gary Galbreath, a biologist at Chicago's Field Museum in Illinois who concluded that the kouprey was a hybrid, told *CBS News*: "It is surely desirable not to waste time and money trying to locate or conserve a domestic breed gone wild." He based that conclusion on the observation that kouprey and banteng (*Bos javanicus*) shared several sequences of mitochondrial DNA.

Now, Alexandre Hassanin and Anne Ropiquet of the National Natural History Museum in Paris have sequenced three regions of mitochondrial DNA and five of non-coding nuclear DNA from seven related species, including kouprey. The pair found that kouprey have unique sequences of both mitochondrial and nuclear DNA. Their data suggest that kouprey should indeed be a conservation priority — if anyone can find one. ■

Ewen Callaway

1. Hassanin, A. & Ropiquet, A. *Proc. R. Soc. B* doi:10.1098/rspb.2007.0830 (2007).
2. Galbreath, G. J., Mordacq, J. C. & Weiler, F. H. *J. Zool.* **270,** 561–564 (2006).

BLICKWINKEL/ALAMY

Cambodia's national emblem, the kouprey, is a distinct species of ox.

A reanalysis of research carried out at the Pasteur Institute casts doubt on a respected hypothesis.

L. BORGHI

Long-held theory is in danger of losing its nerve

A suite of seminal neuroscience papers by Henri Korn of the Pasteur Institute in Paris allegedly contains a string of anomalies in data interpretation, according to a reanalysis of the papers, published this week in the *Journal of Neurophysiology*[1]. But Korn and his co-authors contest this and are critical of the reanalysis, which appears in the same journal as many of Korn's original papers.

The papers, published over the past 25 years by Korn and his co-workers, including Donald Faber of the Albert Einstein College of Medicine in New York, concern the dynamics of the release of neurotransmitter chemicals at the synapse — the junction between nerve cells (see 'Theory of neurotransmitter release moves on'). They suggest that a single bouton (nerve terminus) releases only one quantum of transmitter per nerve impulse. This influential theory has major functional implications, but remains controversial.

A key finding underpinning their theory was based on electrophysiological studies of giant nerve cells in goldfish, called Mauthner cells. Korn and Faber claimed that the number of synaptic boutons counted by light microscopy was highly correlated with the number worked out from an analysis of the amplitudes of the electrical spikes triggered by the neurotransmitter[2]. But their graph of the correlation, with data points lying on a nearly perfectly straight line, is "almost miraculous" given the noise and uncertainties in the underlying data, claims Jacques Ninio, a bioinformatician at the Ecole Normale Supérieure in Paris, who carried out the reanalysis.

Ninio extracted the data from graphs in the papers and recomputed them. "Several theoretical curves were simply not what Korn and co-workers claimed them to be," he says.

Ninio's conclusions add to similar allegations by two researchers who worked in Korn's laboratory — Nicole Ropert, now at the University of Paris Descartes, and Luca Turin, a former researcher for the CNRS, France's basic-research agency, now at University College

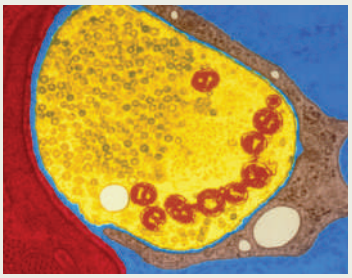
Theory of neurotransmitter release moves on

Neurobiologist Henri Korn at the Pasteur Institute in Paris published a series of papers on the dynamics of synapses (the junctions between nerve cells, pictured) that helped to establish a new field.

Korn and his co-workers' major claim was that each synaptic bouton (the terminus of a nerve cell) releases just one vesicle at a time, each containing a fairly constant number of neurotransmitter molecules. This 'one-release site/one-vesicle' theory had a string of major functional implications for synapse dynamics.

A reanalysis of their papers by Jacques Ninio, a bioinformatician at the Ecole Normale Supérieure in Paris, now questions their results.

Yet much of the research that seems to confirm single-vesicle release itself has problems, say experts. Multiple-vesicle release has been observed in many systems, so univesicular release is at best limited to certain synapses and activities, they add. More broadly, models of both single and multiple release are now considered too simplistic, and there is growing recognition of the great functional heterogeneity among synapses.



D. FAWCETT/SPL

"Nothing that I know of has emerged in the 25 years or so since the analyses under question to confirm that such uniformity of discharge exists at the single-bouton level," says one US neuroscientist, who wishes to remain anonymous. The focus now is on the molecular mechanisms of transmitter release, and statistical analyses can say little about this," he adds.

The field has generally moved on, but several groups still cling to such techniques, agrees John Clements, a former neuroscientist now with Sydney-based AxoGraph Scientific, a data-analysis software firm. "Ninio's paper may help cement the paradigm shift by highlighting and cleaning out some of the historical mess." **D.B.**

London. In 2004, Ropert submitted a 25-page report to the Pasteur Institute's research integrity committee detailing allegations of "events contrary to scientific ethics". But the committee last year opted to close the matter without an independent investigation. A similar request made by Turin to the CNRS in 1989 was also not taken forward.

Ninio's challenge is dismissed in an accompanying response[3] by Korn, Faber and statistician Alain Mallet of the Pierre and Marie Curie University in Paris, who was a co-author on several of the papers. They criticize Ninio's approach, describing it as "qualitative assessments of second-order representations of the data".

"Although we may have made some mistakes — inherent in any scientific inquiry — none of the putative errors invalidates the major findings in our papers," they write. The research, they write, "changed the nature of the scientific discussion about structure–function correlations at synapses". They also add: "Subsequent research carried out independently by a number of eminent scientists supported our proposal of the 'One-Vesicle Hypothesis'."

Nature has obtained the referees' reports on Ninio's paper. One referee comments that Ninio "demonstrates convincingly" that claims in some of Korn and his colleagues' papers are unsuppported. They "are at best erroneous, and at worst deliberate falsifications of the results of the mathematical analysis", the referee alleges. The second referee's report argues that Ninio raises "a disquieting number of discrepancies" and that Ninio "ventures to say what many experienced observers have politely evaded: that at least one of the emperors of French neuroscience has no clothes".

"The 'sheep' mentality is alive and well even at the summits of neuroscience."

Nature put these allegations and referees' comments to Korn, who says he answered the scientific queries raised by Ninio in his published response.

But that rebuttal is "unconvincing, though artful", claims Paul Adams, a neurobiologist at Stony Brook University in New York. "Ninio did the best he could in view of the fact that he did not have access to the original data." Adams describes the Ninio paper as "very useful", saying that published discussions of this issue have not been as sceptical as they should have been. "The 'sheep' mentality is alive and well even at the summits of neuroscience," he says. ■

Declan Butler

1. Ninio, J. *J. Neurophysiol.* **98,** 1827–1835 (2007).
2. Korn, H. *et al. Science* **213,** 898–901 (1981).
3. Mallet, A. *et al. J. Neurophysiol.* **98,** 1836–1840 (2007).

ON THE RECORD

"The American people, our friends, and our potential adversaries must be confident that the highest standards are in place when it comes to our nuclear arsenal."

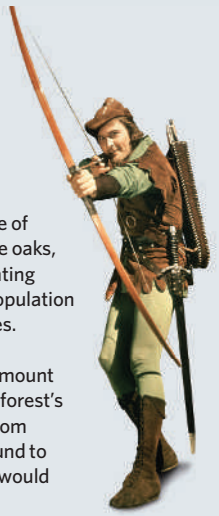
Ike Skelton, chair of the House Armed Services Committee, reacts to the news that B52 bombers accidentally flew nuclear warheads over several states last week with some comforting, yet also strangely chilling, words.

NUMBER CRUNCH

5 is the average number of oak trees now succumbing to old age each year in Britain's Sherwood Forest, home of the fabled Robin Hood.

300 is the minimum age of most of the forest's mature oaks, because of a hiatus in planting that has left the forest's population skewed towards older trees.

£50 million is the amount (US$100 million) that the forest's managers are asking for from Britain's national lottery fund to plant more trees — Robin would surely have approved.



HULTON ARCHIVE/GETTY

SCORECARD

▲ **Smile recognition**
Japanese companies Omron and Sony have both developed camera software that can recognize smiling and laughing faces.

▼ **Video games**
A rare breed of beetle is in danger of being wiped out from its habitat in Turkey after a video game called *Mushiking* ('Insect King') sparked a Japanese craze for the real-life version.

WORDWATCH

Dino-Opoly
With Christmas fast approaching (sort of), a reworking of the classic board game Monopoly aims to combine the thrill of palaeontology with the magic of capitalism.

Sources: Forbes.com, Reuters, The Japan Times, AFP, The Times, LiveScienceStore.com

SIDELINES

Q&A

Interferon discovery and ferret flu

Fifty years ago virologists were struggling to understand why an inactivated virus reduced the ability of a normal virus to infect cells — a process called interference. Jean Lindenmann and Alick Isaacs at the National Institute for Medical Research (NIMR) at Mill Hill, London, found the answer in less than a year of intensive and inspired research[1,2]. The inactivated virus triggered the infected cells to produce a protein that suppressed replication of the live virus. The protein, called interferon, turned out to be a useful therapy for hepatitis C and many cancers. **Jean Lindenmann**, now 83, talks to *Nature*.

How did you come up with the name interferon?

I started out studying physics at the University of Zurich, but 18 months into my course the atom bomb was dropped on Hiroshima. This so disenchanted me that I switched to medicine. I guess I was still jealous of all those elementary particles the physicists had to play with — baryons, neutrons and so on — so I created an '–on' suffix for the viral 'interfering' substance we were looking for, and made interferon. At first it was a bit of a joke name, a sort of slang in the lab, but then it stuck.

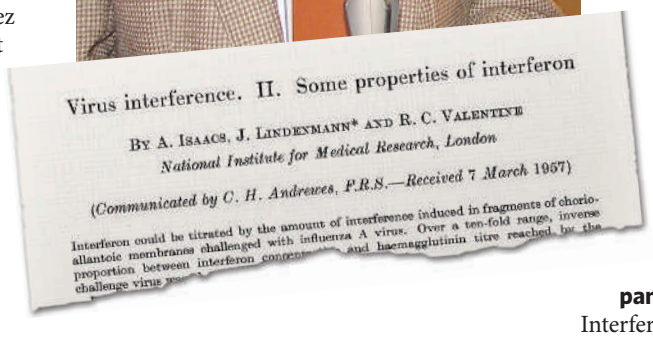
How was working at the NIMR in 1956?

It was informal and very intellectual. Isaacs would whistle arias and we'd all have to guess from which opera they came. There was lots of political discussion. It was the time of the Hungarian uprising and the Suez crisis. A petition was circulated to protest the British invasion of Egypt that had nationalized the Suez Canal. But I felt that I shouldn't sign it because Switzerland was a neutral country.

How did you and Isaacs develop your key experiments?

We adapted a standard experimental model for infection studies that uses the chorioallantoic membrane — the membrane surrounding the growing embryo — in fertilized chicken eggs. The amount of virus generated by infected membranes after each experiment was measured by a bioassay based on the ability of viruses to cause clumping of red blood cells. We used the influenza A virus. Instead of infecting the membrane by injecting the virus into the cavity between the membrane that lines the shell and the embryo — as others did — we peeled the membrane away and cut it into six or eight pieces. We placed each piece in a test tube with nutrient medium. This gave us more experimental values per egg and more experimental flexibility. For instance, we could wash the fragments after each exposure and re-expose them to inactivated or live virus at different times, in different conditions.

A. ABBOTT



Virus interference. II. Some properties of interferon

By A. Isaacs, J. Lindenmann* and R. C. Valentine

National Institute for Medical Research, London

(*Communicated by C. H. Andrewes, F.R.S.—Received 7 March 1957*)

Interferon could be titrated by the amount of interference induced in fragments of chorio-allantoic membranes challenged with influenza A virus. Over a ten-fold range, inverse proportion between interferon concen[illegible] and haemagglutinin titre reached by the challenge virus

ROYAL SOCIETY

Jean Lindenmann collects the European Virology award on 2 September in Nuremberg, Germany.

Is that why you succeeded where others failed?

The director of NIMR at the time, Christopher H. Andrewes, had considered the possibility that interference worked by inducing infected cells to produce an antiviral molecule back in 1942, but he was unable to see it in his tissue-culture experiments. Back then, tissue culture was tricky — after two or three days everything would be contaminated with bacteria. It was easier for us to see the phenomenon in tissue culture in the 1950s because we had access to antibiotics such as penicillin and streptomycin. I often reflect that if the Asian flu pandemic had happened in 1956 instead of 1957, we would not have been able to do this work. NIMR virologists would have been completely occupied with public-health-related work.

Was biosecurity strict in the NIMR then?

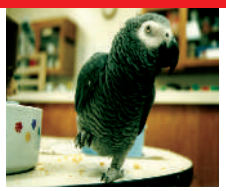
Even with freshly isolated strains we took only the ordinary precautions. The first isolation of an influenza virus by Andrewes and Wilson Smith, done by infecting ferrets intranasally, resulted in a lab infection: Andrewes had fallen ill with naturally acquired influenza. Using his nasal washings, they inoculated ferrets. The fever that the animals developed could be passed on to other ferrets. In the course of such a passage, a ferret sneezed into the face of Smith, who fell ill with influenza; from this lab infection the virus strain now known as WS was established. Andrewes jokingly mentioned that actually this strain ought to be called CHA.

Are you disappointed that interferon never became the panacea it was tipped to be?

Interferon was hyped shamelessly at some phases of scientific research. Of course it never became a miracle cure for cancer and viral infections. Scientists were not innocent in this hyping — it was a way of getting research money. Back in 1957 we were, of course, already thinking about applications, and we patented it. Isaacs enrolled a biochemist to purify and chemically characterize our interferon, expecting it to take six months. In the event it took more than 20 years. ■

Interview by Alison Abbott.

1. Isaacs, A. & Lindenmann, J. *Proc. R. Soc. Lond. B* **147,** 258–267 (1957).
2. Isaacs, A., Lindenmann, J. & Valentine, R. C. *Proc. R. Soc. Lond. B* **147,** 268–273 (1957).



D. CHANDLER

FAREWELL TO A FAMOUS PARROT
Alex, who could talk and count, dies at 31.
www.nature.com/news

Gorillas on the list

The plight of wild gorillas has taken a turn for the worse, according to the latest edition of the World Conservation Union (IUCN) Red List of Threatened Species. Western gorillas (*Gorilla gorilla*), which live in the western Congo basin, have moved from 'endangered' to 'critically endangered' in the 2007 list. And conservationists anticipate that the mountain gorilla (*G. beringei*), which is now found only in Rwanda, Uganda and the eastern Democratic Republic of the Congo, will follow suit once its population survey is completed.

The number of western gorillas has declined by more than 60% in the past 25 years, according to the new assessment, published on 12 September. The Ebola outbreak that has hit the main subspecies, the western lowland gorilla (*G. gorilla gorilla*, pictured) is largely responsible — wiping out roughly one-third of individuals in protected areas.

The upgraded conservation status is mainly the result of the Ebola outbreak and a resurgent trade in bushmeat, says Russ Mittermeier, chair of the IUCN Primate Specialist Group in Arlington, Virginia. "The decline has been really precipitous," he says. "Gorillas are still being sold as a luxury food item."

M. WATSON/WWW.ARDEA.COM

The population of western gorillas has been hit hard by the Ebola virus and poaching.

Mountain gorillas are less numerous than western gorillas, but have not yet been upgraded to critically endangered. This is mainly because the most vulnerable subspecies (*G. beringei beringei*) — numbering barely 700 gorillas in the Virunga mountain range on the border of Rwanda and the Democratic Republic of the Congo — has been kept stable by a popular and well-managed tourism programme. But civil strife in the Congo means that this security is under threat. Last week, the conservation group WWF reported that park rangers had been leaving their posts in fear of armed rebels, who have already killed several gorillas within the Congo's parks this year.

The Red List contains details of some 41,000 species, of which more than 16,000 are officially threatened with extinction. Additions to the 2007 list include three species of coral, the first corals ever to be included. The Baji, or Yangtze river dolphin (*Lipotes vexillifer*), which is subject to widespread media speculation over its status, is now listed as 'critically endangered (possibly extinct)'. ■

Michael Hopkin

Virus linked to puzzling loss of US honeybees

US beekeepers at last have a suspect for the agent behind the mysterious disappearance of worker honeybees across the country.

The loss of worker bees is known as colony collapse disorder, or CCD. A metagenomics study has now revealed that nearly all of the US colonies suffering CCD contained Israeli acute paralysis virus (D. L. Cox-Foster *et al. Science* doi:10.1126/science.1146498; 2007).

The researchers stress that this doesn't prove that the virus is causing CCD — it is possible that the virus is infecting bees while they are in a vulnerable state. But they say they have found a good suspect. "The real test will be introducing the virus to healthy bees," says Ian Lipkin of Columbia University in New York, a co-author on the study.

Damaged pipe caused foot-and-mouth scare

A leaky pipe was the probable cause of last month's outbreak of foot-and-mouth disease in the United Kingdom (see *Nature* 448, 732–733; 2007), say two reports commissioned by the UK government.

The reports found that drainage problems had been observed for years in the plumbing between the government's Institute for Animal Health in Surrey and a nearby private facility that researches vaccines for foot-and-mouth disease. Heavy rain this summer probably led to the virus leaking out, the investigators say. Vehicles associated with nearby construction work probably carried it away from the site.

Neither the institute nor Merial Animal Health, which runs the commercial facility, had been willing to pay to replace the old pipe, according to environment secretary Hilary Benn. Benn says the investigators declined to blame either facility for the maintenance lapses. "Ultimately," he told reporters, "it's a legal question as to who exactly was responsible."

A. DENNIS/AFP/GETTY

Culled livestock is disposed of during the recent foot-and-mouth outbreak in southeast England.

Texas A&M under fire for biosafety shortcomings

US government inspectors have uncovered more violations of safety and security regulations in the bioterror research programme at Texas A&M University in College Station. All such research was halted on 30 June, after it came to light that the university failed to report that workers had been exposed to two possible bioweapons agents.

The Centers for Disease Control and Prevention in Atlanta, Georgia, outlined the offences in a 31 August letter. Among its concerns were the fact that an A&M researcher couldn't account for three vials of the bacterium *Brucella*, a potential bioweapons agent; and that workers entered restricted labs without clearance or medical screening. The university's 'select-agent' research will remain closed until the lapses are fixed, the letter says.

A&M's vice-president for research has resigned from that position, and the school's head of biosafety has left. The university could face fines of $500,000 or more.

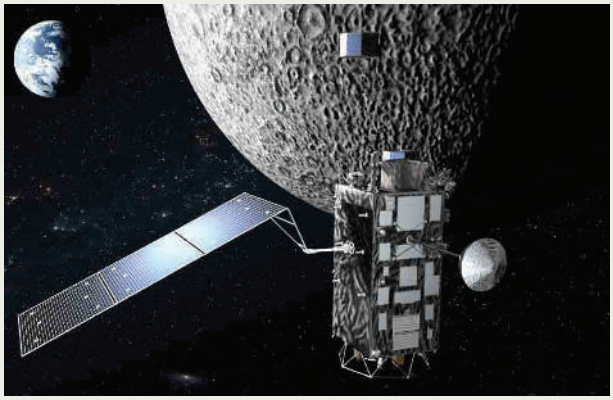
Bubble-fusion allegations merit more investigation

A Purdue University panel inquiring into allegations of research misconduct against nuclear engineer Rusi Taleyarkhan has concluded that "several matters merit further investigation".

The panel was set up after contact between Purdue, based in West Lafayette, Indiana, and the Office of Naval Research, which in 2005 allocated $250,000 to research by Taleyarkhan. This was part of a project aimed at replicating his controversial claims to be able to generate fusion energy by collapsing bubbles in deuterated fluids.

This is the third inquiry run by Purdue, which was criticized earlier this year by both scientists and lawmakers for its handling of concerns raised by scientists about bubble-fusion claims at the university. Purdue has not said publicly what exactly in Taleyarkhan's work it might investigate further.

Selene gears up for trip to the Moon



A. IKESHITA/JAXA

Japan is on the verge of launching its lunar orbiter, Selene (pictured). As *Nature* went to press, the mission was scheduled to rocket into space on 13 September from the Tanegashima Space Center south of Kyushu.

Selene, also known as Kaguya, will map the abundance of chemical elements and minerals on the Moon, and survey features such as gravity and topography. The main satellite will orbit the Moon at an altitude of 100 kilometres, and two smaller satellites will offer relay and radio functions. Selene will be the first mission to be launched to the Moon since Europe's SMART-1 spacecraft in 2003.

Conservationists caught out by wrong type of trout

Efforts to save a near-extinct trout in Colorado have backfired with the discovery that conservationists have been breeding the wrong fish.

Five of nine native populations of the endangered greenback cutthroat trout (*Oncorhynchus clarkii stomias*) that have been bred to restock other areas, were actually a related subspecies, the Colorado River cutthroat trout (*O. clarkii pleuriticus*), according to a genetic analysis (J. L. Metcalf *et al. Mol. Ecol.* doi:10.1111/j.1365-294X.2007.03472.x; 2007). With only four greenback trout populations remaining in a short stretch of creek, the survival of the fish remains seriously threatened, says lead author of the study, Jessica Metcalf of the University of Colorado in Boulder.

Genetic studies may now be used more widely on other recovering species, she says. Government officials had been moving to drop federal protection on the trout.

BUSINESS

A commodity no more

The flat-screen television boom has materials scientists scrambling to replace the valuable metal oxide that coats the screens. **Andrea Chipman** reports.

The headlong rush to flat-screen technology has been warmly welcomed by couch potatoes and office drones alike. But it puts materials scientists on the spot. Can they replace indium tin oxide (ITO) — the material that coats these screens — if overwhelming demand drives its price through the roof?

Demand for indium has skyrocketed in recent years, mainly because of its use in liquid crystal displays (LCDs) and plasma screens. The electrical conductivity and transparency of ITO has turned it into a crucial industrial component, which can be readily etched and patterned to create a thin film of transparent circuits on both sides of the glass screen.

"For years, indium was a niche material, with no major end markets, used in small quantities in hundreds of odd applications," says Brian O'Neill of AIM Specialty Materials, Rhode Island. "In the past ten years it has found one major application, and that is LCDs."

The quandary that this presents is a good example of how high demand for commodities in the booming world economy is forcing materials scientists and engineers to revisit their options. Indium is mined almost entirely as a by-product of zinc — a much more widely used commodity — so little can be done to step up global production in line with demand.

A decade ago, the world was using less than 200 tonnes of indium a year. Now, annual consumption exceeds 1,500 tonnes, according to AIM. Production from mining has grown, but it is still less than 500 tonnes a year and the gap is being filled in the short term by recycling indium, mainly from the scrap produced as ITO films are applied to glass on LCDs.

Expensive element

The price of the soft, silvery-grey metal ballooned from less than US$100 as recently as 2002, to $1,000 a kilogram — more than twice the price of silver — in 2005, but has since subsided to a still-hefty $680 (see graph). That means that the ITO part of a typical flat-panel screen costs about $2, according to O'Neill. In such a fiercely competitive business, that's a substantial cost.

Uncertainty in supply is bolstered by the fact that most zinc deposits are found in places such as Bolivia and China, where exports of the metal are already restricted. And large oscillations in the price deter mining companies from investing heavily in processes to recover traces of indium from the zinc ore.

S. YONG/REUTERS

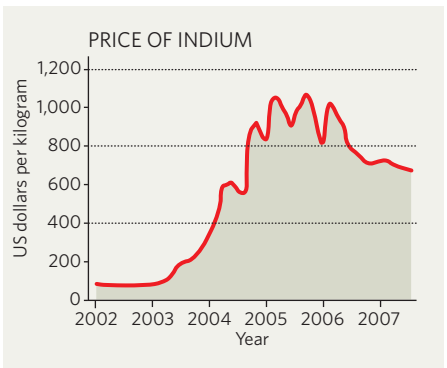
Industry demand has inflated the price of indium.

At the same time, with overall demand for the metal forecast by AIM to rise nearly 60% by 2009, even a projected increase in recycling won't bridge the gap between supply and demand for long. So researchers are testing a range of substitutes for ITO.

Fluorinated tin oxide, used in the doors of supermarket freezer compartments, provides a low-cost method for applying coatings to glass. It has comparable electrical properties to ITO, but is harder and takes much longer to etch, restricting its use in displays. Cadmium tin oxide has the right electrical and physical properties, but is highly toxic.

Zinc aluminium oxide has been cited as a promising substitute, but it needs a thicker film to achieve the same electrical conductivity, says O'Neill, leading to a loss of transparency — and screen brightness.

Researchers are also pursuing solutions that could become viable in the longer term. In Japan — home to many of the largest flat-screen producers — a group from the Japan Science and Technology Agency in Kawasaki is developing a transparent electrical insulating material made from calcium and aluminium oxides that becomes conductive when exposed to ultraviolet light. The material would mirror the properties of ITO, but would be cheaper to make and simpler to pattern, the researchers say. Another potential alternative is carbon nanotubes, which are more durable and flexible than ITO, although they haven't yet been tested on a large scale.



SOURCE: AIM

But even the best substitute will create problems. "You find very complex chemical compounds interacting in an electronic device, such as an LCD display," says Armin Reller, a chemist at the University of Augsburg in Germany. "If you replace one thing you have to adjust the other components. You can develop any system in the lab, but to implement it in a functioning device is not easy."

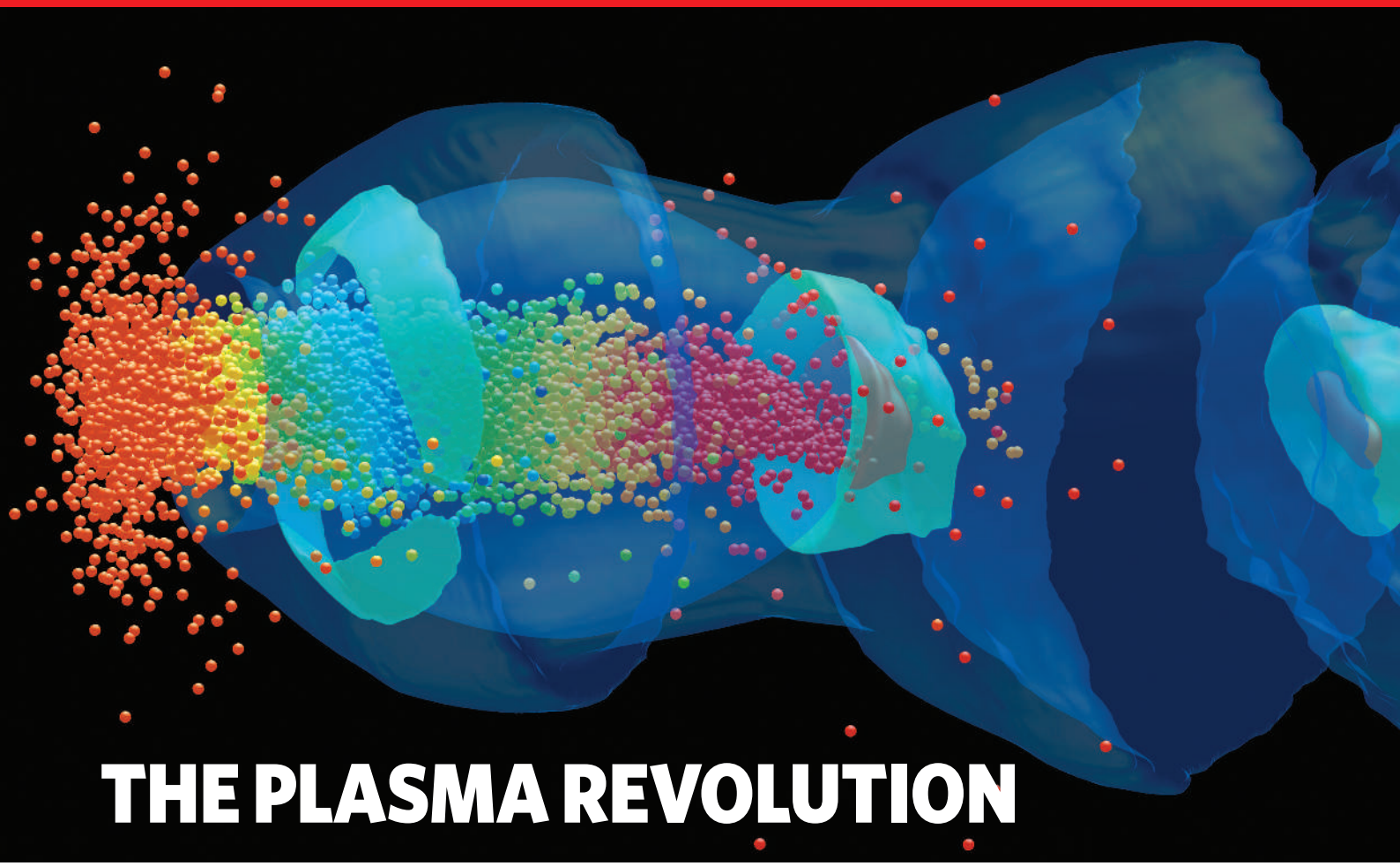
And, O'Neill notes, even a technically viable candidate will have to overcome the fact that display manufacturers have already invested billions of dollars in production equipment dedicated to applying ITO.

Solar power

Further market pressure looms on the horizon. As the second-largest consumer of indium after flat panels, photovoltaics already uses some 20 tonnes of indium a year. But according to O'Neil, that could rise to 150 tonnes a year by 2010 if growth continues at the current rate.

Here, too, materials scientists are working on alternatives. Copper indium gallium diselenide, better known as CIGS, is the main component in one of the leading solar cells under development by the US Department of Energy. Although the approach is less efficient than other solar cells, such as those based on silicon, it is vaunted as cheaper to build. But according to Tim Coutts, a research fellow at the National Renewable Energy Laboratory in Golden, Colorado, 50 tonnes of indium would be needed to produce cells to generate a gigawatt of solar power — the broad equivalent of a large power station.

At the same time, photovoltaics could ultimately be much more vulnerable to the price volatility of indium than the display business is, because it needs so much of the element. The flat-panel industry, Coutts notes, "can afford to pay if the price goes up. In photovoltaics, with increasing prices, it could become non-competitive." ■



THE PLASMA REVOLUTION

Particle accelerators that use plasma technology promise to shake up the fields of high-energy particle physics and cancer treatment. Challenges remain, but smaller, cheaper machines are within reach. **Navroz Patel** reports.

Beyond the theoretical and engineering challenges of building particle accelerators, sheer cost is a concern for physicists whose work involves accelerating and smashing subatomic particles together at great speed. Many particle physicists think that if the planned International Linear Collider — a US$7-billion electron–positron collider that could begin operation within a decade — gets the go ahead, it may be the last large accelerator to be built for many decades as governments put a squeeze on funding.

The cost of accelerators is a concern not just for those who crave bigger and bigger machines to probe ever higher energy scales. Some oncologists think that proton beams could offer superior results to conventional X-ray treatment of some tumours, yet they say the size and cost of the accelerators has limited the number of studies into their clinical effectiveness.

"If we can reduce an accelerator's size, we can reduce the cost of proton therapy to something very small," says Charlie Ma, director of radiation physics at the Fox Chase Cancer Center in Philadelphia, Pennsylvania. Building a proton-treatment centre with conventional cyclotron or synchrotron accelerators costs between $100 million and $200 million, which explains why there are so few of these facilities (see 'Targeting tumours').

But if accelerator research continues to progress at the rapid rate seen in recent years, the economics could be about to change for the better. A handful of groups are working on a new way to accelerate particles — known as wakefield acceleration — that should not only help push physicists towards the next energy frontier, but also provide affordable, table-top accelerators that could revolutionize cancer treatment.

"Experiments over the next few years could make or break our field."
— Wim Leemans

The technique involves passing either a laser beam or a beam of particles through a plasma. The beam scatters electrons, causing an uneven distribution of charge between the scattered particles and the plasma ions. To restore an even distribution, the electrons are pulled back towards the positive plasma ions that have congregated towards the rear of the beam pulse. But the electrons overshoot their original positions, creating a wake-like disturbance called a wakefield oscillation. Within this wake are pockets of plasma ions, which physicists refer to as bubbles, thanks to their spherical shape.

M. ZHOU/F. TSUNG

The wake of a breaking wave causes turbulence, and the wake generated in a plasma is no exception. But as surfers and boat owners know, if you hit the wave at just the right spot, you can be accelerated by its surf. So some electrons can surf the plasma wakefield, as can other particles, such as protons, injected into the beam, accelerating them to very high energies.

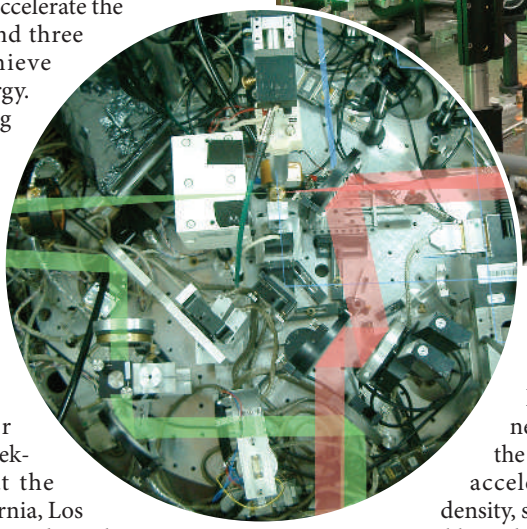
When particle beams are used to create the wake, it is often simply referred to as 'plasma wakefield acceleration', and the disturbance is created through electromagnetic repulsion between the beam and plasma electrons. For laser wakefield acceleration, the radiation pressure from the laser beam causes the wake formation.

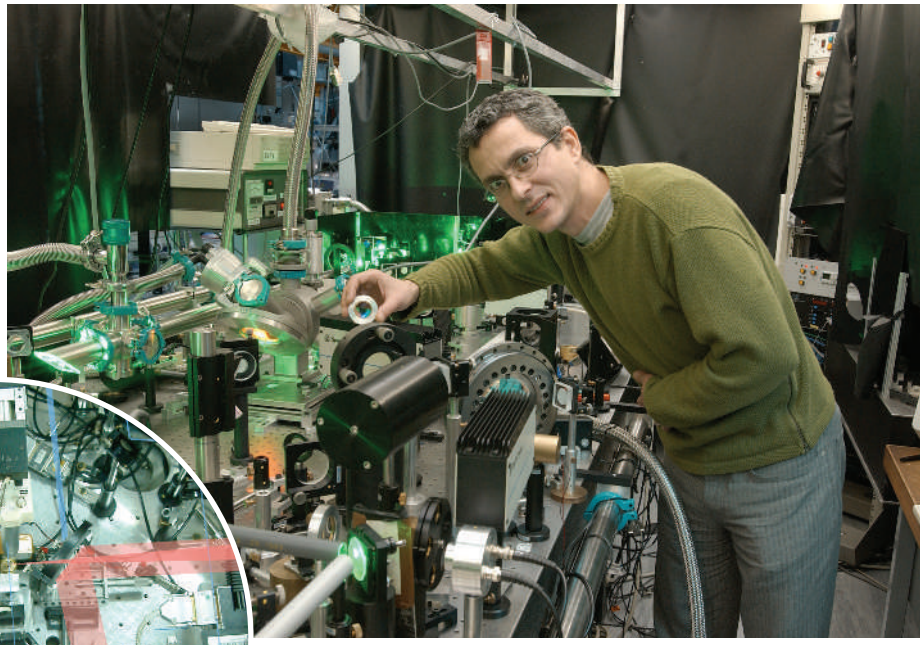
Bubble effect

In the past three years, wakefield acceleration has generated its own bubble of excitement. Swapan Chattopadhyay, director of the Cockcroft Institute, a collaborative accelerator-research centre opened last year in Warrington, UK, says that a wakefield experiment

at the Stanford Linear Accelerator Center (SLAC) in California this year has opened up a new chapter in accelerator physics.

Using a 400-metre extension of the 3.2-kilometre main accelerator at SLAC — the longest linear accelerator in the world — researchers have managed to double the energy of the electron beam over a distance of just 85 centimetres[1]. Much of the beam loses energy in setting up the plasma wakefield, but a few (just 0.02%) of the electrons were accelerated from 42 gigaelectronvolts to around 85 gigaelectronvolts. Conventional technology would have to accelerate the electrons for around three kilometres to achieve this pick-up in energy. "This trick of sending the SLAC's electron beam through a plasma jet to double its energy without having to double the size of the facility is truly remarkable," says Chattopadhyay.





Victor Malka uses a counterposing laser (inset) to control the injection of electrons into plasma fields.

One of the SLAC team, accelerator physicist Chandrashekhar Joshi based at the University of California, Los Angeles (UCLA), says that taking laser wakefield accelerator research to SLAC was the logical next step for the field. "Short-pulse lasers are powerful, but beams typically contain energies of tens of joules," he explains. The energies of particle beams, on the other hand, are of the order of kilojoules. In other words, particle-beam technology can reach much higher energies than contemporary reliable laser technology.

Splitting ions

In theory, there is no limit to the energies that plasma wakefield accelerators could reach. In conventional accelerators, particles are accelerated by an electric field — the steeper the electric gradient, the greater the acceleration. But the field can only increase so far before the surrounding cavity material, such as copper or a superconducting material, starts to break down as electrons are stripped from its atoms. Because plasma, although electrically neutral overall, is already broken down into its atoms and electrons, it can support much stronger electric fields.

The SLAC experiment was a breakthrough on several fronts. It showed that the technology can work at larger distances — reaching almost a metre, rather than the couple of centimetres previously achieved with laser technology. It also produced enough energy to be of interest in high-energy particle physics. But the energy of the accelerated electrons and the distance over which they continue to accelerate are not the only important properties of an accelerator. Other key factors also need to be addressed: the number of particles accelerated, or energy density, should be as high as possible, and the particles need to have a low energy spread, which means that they all have similar energies. With an energy spread of 100%, the SLAC experiment still has some way to go.

Experiments with laser wakefield accelerators, although operating at lower energies and over shorter distances than plasma accelerators, are making progress with these key factors. In 2004, three groups used lasers to accelerate electrons so that they had similar energies and reasonable energy densities, exceeding 10^9 electrons per beam. These experiments reinvigorated interest in wakefield acceleration, which was first proposed[2] by physicists Toshiki Tajima and John Dawson at UCLA a quarter of a century earlier.

But to do particle collision experiments, such as those at SLAC, the beams need to reach energy densities of 10^{34} particles. The tiny fraction of electrons accelerated at SLAC is nowhere near enough for a collision experiment.

Late last year, researchers took wakefield acceleration a step further. The 2004 experiments had accelerated electrons over the 0.1 gigaelectronvolt range, but a collaboration between researchers at the Lawrence Berkeley National Laboratory in California and a team led by the University of Oxford's Simon Hooker in Britain has now boosted electrons to more than 1 gigaelectronvolt[3].

Small steps

This is not yet the high-energy frontier, which sits in the region of teraelectronvolts and beyond, but it is still a respectable gain on earlier experiments. "Our next goal is to go up to 10 gigaelectronvolts, for which we will need a bigger laser — around one terawatt," says Wim Leemans, head of the group at Lawrence Berkeley National Laboratory.

What's more, the researchers were able to create narrower particle beams with tight beam spreads — the energy spread divided by the peak energy. Tight spreads are essential in cancer treatment, as the energy determines how deeply the protons will deposit their maximum energy in the body. The researchers achieved a beam spread of less than 5%, compared with 10% in 2004 and 100% just a few years earlier. But there's still room for improvement. Karl Krushelnick, a wakefield accelerator physicist at the University of Michigan in Ann Arbor says: "For many processes that we would like to use these electron beams for, this figure needs to be well below 1%."

J. FAN

"If we can reduce an accelerator's size, we can reduce the cost of proton therapy to something very small." — Charlie Ma

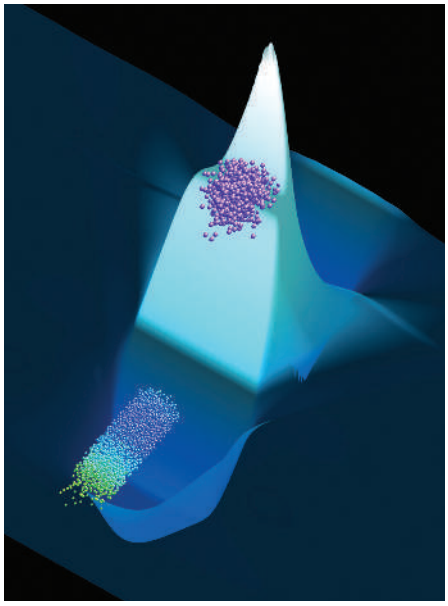
Also last year, Victor Malka and his team at the Ecole Polytechnique in Palaiseau outside Paris developed a technique that uses a second counterposing laser beam to create an electron beam that can have its energy changed

on the fly[4]. The second laser beam is used to control the injection of the electrons that surf the wakefield. The resulting accelerated electrons had an energy spread of less than 10%, and by changing the way that the two lasers overlap the researchers could tune the energy of the beam from 15 megaelectronvolts to 250 megaelectronvolts. Importantly, the beam was much less prone to fail than in previous experimental set-ups.

Particles to the people

"We now have a good understanding and much of the science worked out," says Malka. "In a sense, what we are left with is the technological work needed to improve and stabilize the machines to create a commercial product." The commercial application that Malka has in mind for his group's research is cancer treament. Since 2004, he has been collaborating with a group led by oncologist Uwe Oelfke at the German Cancer Research Center in Heidelberg to perform rigorous simulations comparing proton therapy with X-ray therapy for targeting tumours[5]. The team hopes to apply its results to patients within the next 5 years.

If wakefield researchers make the advances they hope to over the coming years, then table-top accelerators could become much more powerful than they are now. Many experiments that are currently the preserve of relatively few, typically large and costly, facilities will be carried out in the basements of universities using compact and cheap technology. "Experiments over the next few years could make or break our field," says Leemans. "Still, I'm hopeful that we will be able to further address issues such as beam quality and that wakefield acceleration will really prosper."



C. K. HUANG/F. TSUNG

Particles can surf along giant plasma waves.

Even at the high-energy frontier, the next generation of very large accelerators will probably incorporate plasma. According to Krushelnick, plasma wakefields are the only affordable way to achieve the very large acceleration gradients needed to get to extremely high energies, perhaps even the terascale. Plasma techniques may initially be used to boost existing accelerator technology, as with the SLAC experiment, or in the staging of multiple modules to build a plasma wakefield accelerator from scratch. The SLAC team is already trying to work out how numerous small plasma accelerators can be combined to create a reliable machine. And Joshi says that he hopes that he and his team can address all the remaining critical scientific issues and propose an accelerator that is entirely based on plasma within 10 years. ■

Navroz Patel is a writer based in New York City.

1. Blumenfeld, I. *et al. Nature* **445,** 741–744 (2007).
2. Tajima, T. & Dawson, J. M. *Phys. Rev. Lett.* **43,** 267–270 (1979).
3. Leemans, W. P. *et al. Nature Phys.* **2,** 696–699 (2006).
4. Faure, J. *et al. Nature* **444,** 737–739 (2006).
5. Glinec, Y. *et al. Med. Phys.* **33,** 155–162 (2006).
6. Slater, J. D. *et al. Int. J. Radiat. Oncol. Biol. Phys.* **59,** 348–352 (2004).

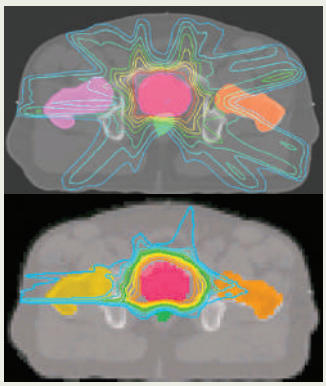
Targeting tumours

Standard radiotherapy can restrict tumour growth in many patients with cancer. It works by delivering high doses of X-rays into the body so that enough molecules are ionized to damage tumour cells. But because they are difficult to target precisely, X-rays often damage healthy tissue around the tumour, so doctors cannot use as high doses as they would like.

Proponents of proton therapy argue that the protons in a particle beam should be able to target tumours more precisely than X-rays do. This is because protons lose most of their energy just before coming to a standstill when travelling through matter. Maximum ionization will thus occur as the protons approach their targeted stopping point, which depends on the energy of the beam, leaving healthy tissue largely untouched. Computer simulations performed by oncologist Charlie Ma and his colleagues at the Fox Chase Cancer Center in Philadelphia support the idea that proton beams generated by wakefield accelerators can target tumours much more accurately than conventional radiotherapy techniques (see simulations, right).

Others argue that the radiation biology of proton therapy is poorly understood and the claimed superiority of particle beams over conventional radiotherapy has not been demonstrated sufficiently in the clinic. "Proton beams have favourable physical characteristics, but the question is: will that translate to improved clinical outcomes?" asks Steve Hahn, a radiation oncologist at the University of Pennsylvania's School of Medicine in Philadelphia. "Answering that is probably going to take randomized phase-III trials."

Ma says that he has some sympathy for this view but argues that costs have limited the acceptance of proton therapy, since it was first proposed in the 1940s. Existing proton-therapy machines use large and expensive conventional accelerators, and so need a lot of space. The radiation shielding alone can cost around US$40 million, according to Ma, with the total price tag for a proton-treatment centre reaching $100 million or more.



C. MA/FOX CHASE

The radiation dose (coloured lines) can be distributed more tightly around a prostate tumour (red) with proton therapy (bottom) than with conventional radiotherapy.

With so few clinical facilities in the world, phase-III trials of the sort Hahn is asking for have been few and far between. In one of the largest clinical studies[6] reported so far, 1,255 men given proton therapy for prostate cancer had survival rates equal to those for conventional radiotherapy and surgery, but with fewer side effects. Ma thinks that affordable wakefield accelerators offer the best way to address concerns over clinical outcomes. Hahn agrees: "Wakefield acceleration promises to make the technology cheaper and widely available and so should help resolve the empirical controversy."

Ma is hopeful that the laser wakefield facility his group is developing in the lab will soon be converted into a clinical system. If all goes to plan, then the Fox Chase Cancer Center will start treating its first patients in the next 5–10 years, and become a prototype clinical facility for a new generation of compact proton-therapy centres.

In Germany, oncologist Uwe Oelfke at the Cancer Research Center in Heidelberg thinks that he could start using wakefield accelerators on patients with hard-to-treat eye cancers as soon as proton beams of 70 megaelectronvolts are available — some 12 megaelectronvolts more than has been achieved so far with laser wakefield acceleration. "If something like this could be built and operate reliably, it would be a huge step," he says. **N.P.**

NATURE|Vol 449|13 September 2007

Underground networking

Above ground, plants compete for life-giving sunlight, but below the surface a more complex picture emerges. **John Whitfield** explores the role of mycorrhizae in plant ecology.

Under the deep shade of northern forests in North America and Europe, a tiny plant ekes out an existence. *Orthilia secunda*, the serrated wintergreen, is found huddled in the understorey of pines and birches, sending up a drab bundle of yellow-green flowers. Beneath ground, however, this meek plant hides a secret. Starved of sunlight, *Orthilia* has found another source of nourishment. Its roots tap into a network of soil fungi, taking up to half the carbohydrates it needs from organisms that normally form mutually beneficial relations with plants but giving nothing back.

"You could say that the plant is eating the fungus," says Marc-André Selosse of the Centre of Functional Ecology and Evolution in Montpelier, France, one of the team that detected *Orthilia*'s thieving[1]. But *Orthilia* is also, indirectly, eating other plants — probably the very trees that tower over it. After all, it is their photosynthetic efforts that the fungus is feeding on. Indeed, *Orthilia*'s freeloading may reveal the actions of an invisible hand in ecology; that is, the fungal network that underlies the forest and participates in a cooperative give-and-take with hundreds of plants.

Emerging clues suggest that this covert subterranean interplay influences many aspects of the forest community, including which plants live, which die, the effects of physical stresses such as heat and drought, and what happens after the introduction of new species. Add the controversial possibility that fungi mediate resource sharing between different plant species and a picture emerges of a Robin Hood of the soil, subsidizing those less able to make food, and by so doing, helping its own cause by promoting a diverse range of plant partners.

Just how interactive this fungal support system is remains unclear. The key processes happening underground are carried out by microscopic players and are hard to track. Moreover, no one knows how to measure

D. PARKINS

fungal fitness. Is it reflected in the number of mushrooms — spore-bearing, fruiting bodies — above ground, the size of the vast network below, or something else entirely? Even the process by which the plants and fungi exchange nutrients is unknown.

"We know so little that it's possible to propose some very naive hypotheses," says Martin Bidartondo of the Royal Botanical Gardens in Kew, London. Researchers are now looking to fill some of the huge gaps in their knowledge of the basic biology and natural history of fungal networks and to improve understanding of the their ecological consequences. Most can agree that fungal networks are real, and important to the lives of plants, but they disagree on how the effects manifest themselves.

Give or take?

About 80% of land plants have fungi called mycorrhizae growing in and on their roots. The fungi extend throughout the soil matrix; it has been estimated that a single gram of soil can contain 100 metres of mycorrhizal filaments. By vastly expanding root surface area, the fungi help plants extract water and nutrients, most importantly nitrogen and phosphorus, from the soil; they also protect plants against soil pathogens. In return, the plant provides carbohydrates. As much as 20–40% of the products of photosynthesis can flow back into the fungus. Most plants and fungi are promiscuous in their associations, mingling with a range of partner species, and creating the potential for one fungus to link the roots of dozens of different plants. DNA fingerprinting has shown such multiple links with matsutake fungus[2].

But more than 400 species of plant have reneged on this contract. They do not photosynthesize at all and have no chlorophyll. Initially, scientists thought that they were decomposers; in fact they parasitize mycorrhizal fungi. This way of life has evolved many times in different plant groups, but it is particularly common in the orchid, heather and gentian families[3]. The more recent discovery that green plants such as *Orthilia* can also use fungal carbon suggests that the ability may be more widespread and ecologically significant than was once thought. "It was a dogma in botany that green plants were autotrophic, but this is no longer valid," says Gerhard Gebauer of the University of Bayreuth, Germany. A team including Gebauer and Bidartondo recently found five green orchid species that can use mycorrhizal carbon to thrive[4]. "This allows the orchids to move into the deepest shade in the forest," says Gebauer. "They can live as pioneers without any herbaceous competitors."

Bidartondo suspects that all orchids, which have very small, poorly provisioned seeds, take from fungi before they become self-sufficient. This start-up funding could have a huge ecological impact, he says: "If flow from fungi to plants happens when plants are getting established, it can really affect competition, even though the amounts moving might not be massive."

Wood-wide web

That some plants can take advantage of the fungal network is not in dispute — although it is not known how they do it. More contentious is whether the flow of nutrients between plants via fungi is a general and significant feature of the 'wood-wide web' of forest ecosystems. Suzanne Simard of the University of British Columbia in Vancouver, Canada, says fungal networks may allow trees to support their own seedlings, perhaps providing the trees with an evolutionary benefit. "There's lots of evidence that mature trees facilitate the growth of conspecific seedlings beneath them, and the evidence is growing that networks are important," she says. Experiments with oaks, for example, show that acorns planted near their own kind do better than those planted near maples[5], and Simard has found something similar with fir trees. Like Bidartondo, she thinks that these boosts to seedlings are mediated by fungi and are ecologically significant. "Early growth sets the trajectory of the ecosystem," she says.

> **"You could say the plant is eating the fungus."**
> — Marc-André Selosse

Such effects can be seen even between different species. In a study[6] published in *Nature* in 1997, Simard's team provided the leaves of paper birch (*Betula papyrifera*) and Douglas fir (*Pseudotsuga menziesii*) seedlings growing half a metre apart with carbon dioxide containing either carbon-14 or carbon-13 —rather than the usual carbon-12. They could thus detect the element moving in either direction between birch and fir. Move it did; what's more, trees growing in shade received more from the other plants in their network. After 9 days, an average of about 4% of the carbon isotopes given to each plant had shown up in the leaves of the other species, and the amount of carbon flowing from birch to pine doubled when the pine seedlings were shaded.

Although laboratory studies had shown carbon moving between plants via fungi, Simard's study was a watershed in showing that the phenomenon happened in the field. The work demonstrated that carbon could flow both ways, that a significant amount of carbon moved, and that the quantity depended on environmental conditions. Since the work was published, other studies have shown that the timing of plant growth, as well as the light environment, affects the dynamics of transfer, with sugar maple seedlings gaining resources from a fast-growing perennial, the trout lily, in spring, and returning the favour in the autumn[7]. It has also been suggested that one thing that makes spotted knapweed a pernicious invasive weed in the United States is its ability to steal resources, via fungi, from other plants — one study found that as much as 15% of the carbon in the knapweed's shoot came

S. SIMARD

B. HEATH

Fungal network: Suzanne Simard studies carbon transfer between young seedlings.

from Idaho fescue, a native grass[8].

Simard is currently working on the interactions between fungi and Douglas fir at the edge of the tree's range, in the dry regions where forest shades into grassland. In such places, water can also move between trees via fungi, she has found. "Mycorrhizae are more important in more stressful climates," she says. By helping plants cope with stress, and by helping seedlings survive, she thinks that fungal networks make plant communities more stable in the face of environmental stress, and quicker to recover from damage.

By distributing resources between different species, says Simard, fungi can preserve a variety of plant partners and insure against the effects of plant disease or herbivores. "If the fungus can form a bigger or more diverse network, its chances of survival are better," she says. Selosse says he thinks that fungi might help young plants to get established because it helps them compete with other fungi in the soil — nourishing an existing partnership might be a more effective strategy than seeking out new hosts. Alternatively, he suggests, it could be that some plants provide trace amounts of vitamins or even hormones in return for fungal carbon.

Up-rooting claims

But not everyone is convinced. "There's been some wishful thinking, and the evidence hasn't been looked at critically," says David Robinson of the University of Aberdeen, UK. "I don't think there's any convincing evidence for resource sharing between plants by mycorrhizal transfer." Robinson, working with Alistair Fitter of the University of York, has suggested that the carbon might have moved through the soil, rather than the fungi — in Simard's experiments, a small amount of radioactive carbon also showed up in a plant species that did not share mycorrhizae with the fir and birch. And experiments by Robinson, Fitter and other groups have found that, although elements do move between plants via fungi, they stay in the root system, and never make it into leaf and stem, suggesting that the resources are stored in the fungal tissue, and not released to the plant[9]. "The moving carbon is primarily a fungal resource," says Robinson. Fitter adds that, from a darwinian viewpoint, it is "highly implausible" that a plant would benefit from helping its neighbours.

"There's no doubt that carbon moves through the soil," says Simard. "I think it goes through both mycorrhizae and soil." But she believes that evidence for transfer between plants has strengthened over the past decade — her group has recently redone the experiments with birch and fir, for example, and found a similar result. Experiments such as Robinson and Fitter's were conducted using grassland plants — a "quite different system" from woodlands, she says. Simard adds that, rather than simply shunting carbohydrates from one plant to another, the fungus might first use the carbon to make amino acids — nitrogen-containing compounds that form part of the usual carbohydrates-for-nitrogen exchange between plant and fungus. This avoids the problem of explaining why a fungus would want to give back its hard-won nourishment

EYE OF SCIENCE/SPL

Fungi (thin threads) on the surface of plant roots form symbiotic associations with the plant.

Even if resources do not flow from plant to plant, the mycorrhizal network has other ways to influence plant interactions. "You don't need direct resource translocation to have benefit or disadvantage moving between plants via a fungal network," says Minna-Maarit Kytöviita of Oulu University in Finland. She has found that some seedlings do worse when hooked into the mycorrhizal network. In greenhouse experiments using herbaceous plants, seedlings do best with mycorrhizae. But when adult plants are present, seedlings do no better with mycorrhizae than without them. The fungi seem to be making the competition more intense — Kytöviita says they might be supplying more to the adult plant that gives more in return, and withhold favours from the seedling. She has also found that seedlings do better when their adult competitors are defoliated, and so less able to supply their fungi[10].

It's in this spectrum of positive and negative interactions between plant and fungus that we should be seeking the influence of mycorrhizal networks, says John Klironomos of the University of Guelph, Canada. His experiments testing different combinations of plants and fungi have found that outcomes can range from exploitation, to mutualism, to neutrality[11]. A fungus might nourish one plant it links to, exploit another and be cheated by a third. It's the diversity of possible interactions between one fungus and the many plants in its network, not transfer between plants, that is ecologically significant, says Klironomos. "I'm convinced that mycorrhizal networks exist, but I'm not sure the mechanism of action is carbon transfer. What's more exciting is the other resources that the fungus is transferring to different plants, and the different amounts of carbon the fungus demands from plants. When you put all that into the equation you get some interesting dynamics."

"If the fungus can form a bigger or more diverse network, its chances of survival are better."
— Suzanne Simard

Filling in gaps

What is clear is that researchers have their work cut out for some time to come. Studies so far have tended to look at two plant species linked by one fungus — a gross simplification of real-world diversity. Such studies have been snapshots, but fungi and trees can live for centuries. And biologists don't know how the links between plants and fungi affect the survival and reproduction of each party. For fungi, says Bidartondo, we're not even sure how to measure that.

Fitter believes that the priority should be to start filling in the large gaps in the understanding of mycorrhizal fungi. We don't know the extent of fungal diversity, he points out, or of the mechanisms of exchange between plants and fungi. "It seems almost certain that mycorrhizae have a huge importance in biodiversity and a number of ecosystem services. But we're a long way from knowing what that is," Fitter says. "Until we've really got a proper understanding of basic fungal biology, we'll find it difficult to understand the ecological mechanisms." ■

John Whitfield is the author of *In the Beat of a Heart: Life, Energy, and the Unity of Nature*.

1. Tedersoo, L., Pellet, P., Köljalg, U. & Selosse, M.-A. *Oecologia* **151,** 206–217 (2007).
2. Lian, C., Narimatsu, M., Nara, K. & Hogetsu, T. *New Phytol.* **171,** 825–836 (2006).
3. Bidartondo, M. I. *New Phytol.* **167,** 335–352 (2005).
4. Bidartondo, M. I. *et al. Proc. R. Soc. Lond. B* **271,** 1799–1806 (2004).
5. Dickie, I. A., Koide, R. T. & Steiner, K. C. *Ecol. Monogr.* **72,** 505–521 (2002).
6. Simard, S. W. *et al. Nature* **388,** 579–582 (1997).
7. Lerat, S. *et al. Oecologia* **132,** 181–187 (2002).
8. Carey, E. V., Marler, M. J. & Callaway, R. M. *Plant Ecol.* **172,** 133–141 (2004).
9. Pfeffer, P. E., Douds, D. D., Bucking, H., Schwartz, D. P. & Shachar-Hill, Y. *New Phytol.* **163,** 617–627 (2004).
10. Pietikäinen, A. & Kytöviita, M.-M. *J. Ecol.* **95,** 639–647 (2007).
11. Klironomos, J. N. *Ecology* **84,** 2292–2301 (2003).

Cover: choosing the right gecko is a sticky business

SIR — Being a herpetologist, I am excited to see a reptile or amphibian prominently displayed on the cover of *Nature*. Such was the case with the 19 July 2007 cover, featuring a Leopard Gecko (*Eublepharis macularius*) clinging to a mussel.

My excitement was tempered, however, when I realized that the wrong species of gecko had been used to draw readers' attention to a Letter describing a new reversible wet/dry adhesive (H. Lee, B. P. Lee and P. B. Messersmith *Nature* **448,** 338–341; 2007).

This "hybrid biologically inspired adhesive" was developed by combining the adhesive properties of microscopic gecko-footpad hairs with wet adhesive proteins found in mussels. The large size of the Tokay Gecko (*Gekko gecko*) has made this species a model organism for most studies detailing the adhesive properties of the microscopic footpad setae in geckos. However, not all geckos are created equal. One clade of geckos, the Eublepharidae, lack these keratinous hairs (G. Underwood *Proc. Zool. Soc. Lond.* **124,** 469–492; 1954) and, unfortunately, the Leopard Gecko used on the cover is a eublepharid.

The technological advances of this adhesive research were only possible following detailed descriptions of gecko and mussel morphology and physiology. This seemingly trivial case of transposed taxa on the *Nature* cover emphasizes the need for all of us to have a much better grasp of the biology and natural history of the animals we work with, rather than of a small portion, or in this case, a toe.

Travis LaDuc
Texas Natural History Collections, Texas Natural Science Center, The University of Texas at Austin, 10100 Burnet Road, Austin, Texas 78758, USA

Cover story may obscure the plane truth

SIR — Should *Nature* use deceptive photographs for cover illustrations?

On the front of the 2 August 2007 issue, several photographs have been cobbled together to depict "three stacked, autonomous, unmanned aircraft" taking atmospheric measurements. Besides the disagreement in ambient lighting between the clouds and the aircraft, it is clear that the top and bottom craft are the same images, right down to the same flat-bottomed tyres, presumably extracted from a photograph taken on the ground. If these two craft were actually flying in tight formation, they were miraculously caught at the instant they crossed the path of the middle craft flying at an appreciably different angle.

For photographs, scientific journals now go to some length to ensure that what appears within their pages genuinely represents the claims of the authors. With a tad more creativity, eye-catching covers can be made without sacrificing truth in journalism.

Lawrence Sincich
Beckman Vision Center, University of California, San Francisco, 10 Koret Way, San Francisco, California 94143-0730, USA

The cover caption should have made it clear that this was a montage. Apologies — Editor, *Nature*.

Researchers' ethical duties are not to be outsourced

SIR — Your News Feature 'Trial and error', on the problems with research ethics committees designed to establish whether a proposed experiment is ethically sound (*Nature* **448,** 530–532; 2007), presents avoidance of liability and the desire to retain power as the main reasons why institutions favour local control over centralized review. But institutions are ethically, not just legally, responsible for what happens to human subjects under their care.

Research is a suspect activity designed to advance knowledge, not benefit individuals. This does not denigrate its importance but rather reminds us why experiments involving humans are regulated differently from other kinds of research, and more heavily.

If a central institutional review board says it's fine to enrol patients into a project, this does not mean that the institution involved can ignore its obligation to protect the rights and welfare of human subjects in its facility.

Any institution that outsources its ethical responsibilities towards subjects should not be allowed to conduct research on human beings.

Leonard H. Glantz
Department of Health Law, Bioethics and Human Rights, Boston University School of Public Health, Boston, Massachusetts 02118, USA

The Vietnam War added a motive to go on studying

SIR — Tony Dahlen's obituary (*Nature* **448,** 268; 2007) comments that Dahlen "could have graduated early in 1968, but decided to satisfy his broad interests by spending a further year sampling courses in other areas".

Another reason may have been to avoid being drafted into the armed forces and the Vietnam War. The law provided a deferment so long as you remained in an educational programme. On reaching the age of 26, you were excused on the grounds of age.

Back in the 1960s, American men born in the 1940s were well-advised to stay in school. I know, because I should have done so: getting my PhD at 25 in 1969 meant I got drafted.

F. Christian Thompson
Systematic Entomology Laboratory, ARS, USDA, Smithsonian Institution MRC-0169, Washington DC 20013-7012, USA

Starstruck science should appreciate philosophy

SIR — As French researchers who are convinced of the need for university reform, we read with interest your News story on the reform plans of the new French government ('French universities to gain control' *Nature* **448,** 113; 2007). We were surprised, however, that you seem to take for granted that a 'star' biologist ought to earn more than a philosopher of the same seniority level.

Are biologists compared with philosophers because it's assumed that there are no stars in philosophy? Or is philosophy thought to be of less value than biology as an academic endeavour? We are keenly aware of the achievements and promise of biology, but we think it would be counterproductive to relegate philosophy to a secondary status.

Although its contribution is difficult to quantify, philosophy has proven its usefulness to science in several ways: as a source of inspiration and new concepts, as an invaluable critic and as a conduit between scientists and the general public. For example, consider the fertile interplay among several branches of contemporary philosophy and current neuroscience.

There are, of course, good reasons for a state to invest more money in a field such as biology than in philosophy. Indeed, in France, much higher funding for biology is reflected in a larger number of teaching and research positions, dedicated laboratory funding and so on.

But paying 'star' biologists higher salaries is debatable for several reasons — not least because, by the time many scientists are recognized as stars, their period of productivity is largely over.

French universities face numerous problems: gross underfunding, laws against selecting students and detachment from the private sector, to name a few. But we would argue that the creation of a star system, among researchers or among disciplines, is not the most urgent necessity.

Mark Wexler*, Stéphanie Dupouy†
*Laboratory of Perceptual Psychology, CNRS, and Université Paris Descartes, 45 rue des Saints-Pères, 75006 Paris, France
†Department of Philosophy, École Normale Supérieure, 45 rue d'Ulm, 75005 Paris, France

COMMENTARY

Universities and the money fix

Funding woes plague US biomedical researchers. But calls for more funding ignore the structural problems that push universities to produce too many scientists, argues **Brian C. Martinson**.

Federal funding for biomedical research is a substantial investment in the US science community. Earlier this year, representatives of several major research universities testified before Congress and issued a report arguing that the budget of the National Institutes of Health (NIH) in Bethesda, Maryland, is insufficient to sustain "a strong and vibrant program of basic research"[1]. They pointed to stifling of innovation and damage to the career prospects of young scientists, ultimately warning that there could be a threat to US pre-eminence in biomedical research if Congress does not increase levels of funding for the NIH. Yet, what is it that poses the most potent threat to the future of biomedical research — a lack of resources, or our failure to manage the level of competition for available resources? The answer to this question is vital if society is to gain maximum benefit from the public money invested in biomedical research.

B. MELLOR

There is undeniably excessive competition for NIH grants, and we should all be concerned about the negative effects this may have on the robustness of the research engine; by damping scientists' willingness to pursue high-risk projects; by causing them to spend excessive time in pursuit of funding; or by causing talented individuals to shun research careers. Yet, largely because of the structure of the funding flows between the NIH and the universities, there are few checks in the system to keep competition for grant funding at a healthy level. Thus, calls for further increases in the NIH budget may only make matters worse. In my view, it is time to ask the biggest beneficiaries of NIH largesse — the universities and academic health centres — to find ways to balance supply and demand that better reflect their obligations to researchers and society.

University leaders know that when the money gets tight, it's junior faculty members who feel the pinch. They are less established in their careers, more peripheral to their professions and institutions, and often most dependent on obtaining NIH funding as an implicit or explicit condition of their continued employment. As NIH funding becomes harder for junior researchers to obtain, we might expect them to experience the elevated levels of depression, anxiety and job dissatisfaction documented in a survey[2] of medical faculty members in 2006. We might also expect the greatest effects to be felt by female scientists and those from minority groups, for younger researchers to leave science, and to see somewhat less ethical behaviour among those who stay. The robustness of the research engine must be judged on more than the level of funding or the number of scientists.

The doubling of the NIH budget between 1998 and 2003 was intended to increase success rates in obtaining NIH grants[3], which have been declining since the mid-1970s. Yet, the budget rise did not have its intended effect, and by 2003, grant-application success rates were slightly worse than before. What happened? The budgetary increases were swamped by an equally large escalation in the number of NIH applicants and applications (see graph, overleaf)[4]. In 1998, there were about 19,000 scientists applying for competing awards; in 2006 there were approximately 34,000.

Since the 1970s academic researchers in biomedicine and the institutions that employ them have become increasingly dependent on NIH dollars[5]. The financial reasons for this are simple. The 'direct costs' of NIH grants cover the fixed costs of faculty salaries, whereas 'indirect cost recovery', pays for operational overheads, capital equipment and other expenses. Federal training grants also provide revenue streams for doctoral and postdoctoral training, directly stimulating workforce growth. Even before the doubling in funding, the Bayh–Dole act of 1980 created incentives for universities to grow their NIH workforce by permitting employers to own the inventions their employees created with federal funding.

Ageing cash cows

As dependence on NIH grants has grown, they have also become harder to obtain, especially for junior scientists. The average age at which PhD scientists earn their first independent support from the NIH has increased steadily[6], from 34 in 1970 to 42 in 2006. The situation has certainly been made worse by the flat NIH budget (declining after taking inflation into account) since the end of the doubling initiative. Yet, the excessive demand for NIH funds predates the recently flat budget (see graph, overleaf). Since the early 1980s new investigators have been entering NIH funding at a more rapid rate than experienced investigators have been exiting[4], leading to a population increase.

With academic faculty members seen as revenue generators, they are encouraged in subtle and not-so-subtle ways to expend greater effort on lucrative activities: this has made research a preferred activity over teaching or patient care. It also means they must spend a substantial amount of time writing grants. This arrangement generally works in the universities' favour, but the downsides of the dependence on NIH funding are becoming harder to ignore.

For too long now, financial incentives to the universities have been aligned to promote unlimited growth in the number of

biomedical researchers seeking funding from the federal government, despite the realities of finite resources. Some have suggested that a solution lies in biomedical researchers and universities becoming less dependent on NIH money by finding commercial funding sources and philanthropies[7], but this approach is not without its own risks, and it avoids dealing with the structural arrangements that keep us from applying sound principles of supply and demand to the scientific workforce.

We need to look at both the supply and the demand sides of the NIH funding equation. Most who worry about these issues have focused on the size or distribution of the pool of NIH dollars. Far fewer have given consideration to the size or dynamics of the population of biomedical researchers living on NIH funding. Few have overtly asked the question — are there too many biomedical scientists?

There are insufficient 'feedback loops' linking the production of biomedical researchers to the availability of resources to support them. Instead, the educational system is replete with incentives to generate ever more PhDs and medical doctors. In the short term these arrangements may benefit universities, but in the longer term, such extreme levels of competition for funding are unsustainable. And they may already be doing harm. Difficult funding decisions are increasing ill will, perceptions of injustice, and eroding the bases of ethical behaviour among academics. Some of my own work leads me to believe that the current situation may be adversely affecting the integrity of research[8].

The needle and the damage done

The imbalance between the supply of NIH funding and the potentially unlimited demand for grants threatens the future of US biomedical science. I have argued that because of structural incentives, demand for NIH grants is largely a function of the size of the biomedical workforce. Recent NIH initiatives to increase funding of junior researchers are welcome, and have the best chance of maintaining a pool of new research talent. But without some counterbalance, these initiatives may escalate competition for grants.

B. KRIST/CORBIS

Campus overload: are universities producing too many biomedical scientists?

Calls for further increases to the NIH budget are a facile response from institutions overly dependent on NIH dollars. But they are an incomplete, and potentially dangerous, answer to the problems of excessive competition. And although short-term NIH budget increases to make up for inflation-related declines since 2003 seem reasonable, further increases risk fuelling, rather than reducing, demand. For now, budgetary increases that simply keep pace with inflation would seem prudent, so as not to reactivate the growth impulse. Regrettably, the current imbalance may be addressed only through a reduction in the biomedical workforce; something that already seems to be happening.

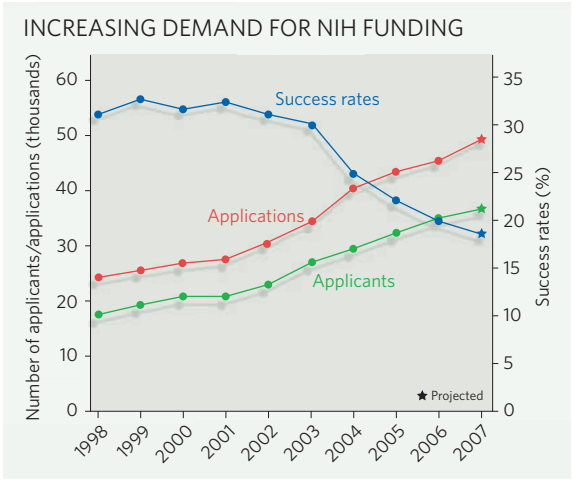
There are two main routes to contraction of the academic workforce today — through tenure failures, and with younger investigators shifting from academia into industry research[8]. This is worrisome for university research in particular because history suggests that the most dramatic innovations come from the young. So is the only solution to force long-time NIH grant getters into retirement? Perhaps not. Universities have benefited handsomely from the efforts of senior faculty members in securing NIH grants during their careers, perhaps those same universities could now return the favour by taking full responsibility for paying these faculty salaries in their later years. This would serve the dual purpose of getting them off the NIH dole, and encouraging them to share their knowledge with their younger colleagues through more teaching.

This won't be easy. Given the levels of dependency on NIH money, it is akin to asking an addict to give up an easy fix. And not all universities will be in financial positions to employ this strategy, but it's difficult to imagine that richer institutions — some of whom acknowledge that their success lies in capturing an increasing share of the NIH pie[9] — could not lead the way in this. Prospective students and their parents may also look favourably on senior faculty members spending more time teaching.

"What is needed is not necessarily more people, but more time, space and freedom."

An implicit assumption underpinning the current system of funding is that having more biomedical scientists automatically leads to greater innovation and more breakthroughs. Yet what is needed is not necessarily more people, but more time, space and freedom for existing researchers to ask questions in new ways, to be willing and able to take risks, and to innovate rather than simply writing safe, incremental grants. The excessive competition for NIH funds discourages this kind of risk-taking, and ultimately reduces opportunities for the sort of creative thinking that leads to major scientific breakthroughs. ■

Brian C. Martinson is at HealthPartners Research Foundation, Minneapolis, Minnesota 55440-1524, USA.



FASEB/NIH/SCIENCE

1. Brugge, J. S. *et al.* Within Our Grasp — or Slipping Away? Assuring a New Era of Scientific and Medical Progress (2007); http://hms.harvard.edu/public/news/nih_funding.pdf
2. Schindler, B. A. *et al. Acad. Med.* **81,** 27–34 (2006).
3. http://grants.nih.gov/grants/award/success/Success_ByIC.cfm
4. Zerhouni, E. A. *Science* **314,** 1088–1090 (2006).
5. Kohn, T. H. (ed.) *Academic Health Centers: Leading change in the 21st Century* (National Academies, Washington DC, 2004).
6. http://opa.faseb.org/pages/PolicyIssues/training_datappt.htm
7. Loscalzo, J. *N. Engl. J. Med.* **354,** 1665–1667 (2006).
8. Martinson, B. C., Anderson, M. S., Crain, A. L. & de Vries, R. *J. Empir. Res. Hum. Res. Ethics* **1,** 51–66 (2006).
9. Bowman, M. A., Rubenstein, A. H. & Levine, A. S. *J. Am. Med. Assoc.* **297,** 2521–2524 (2007).

Acknowledgements The author is supported by NIH funding.

BOOKS & ARTS

A challenge to Kyoto

Standard cost–benefit analysis may not apply to the economics of climate change.

Cool It: The Skeptical Environmentalist's Guide to Global Warming
by Bjorn Lomborg
Knopf/Cyan-Marshall Cavendish: 2007.
272 pp./256 pp. $21/£19.99

Partha Dasgupta

Bjorn Lomborg's *The Skeptical Environmentalist* created a sensation six years ago. The author offered figures to dismiss claims that the ecological-resource base in many parts of the world is deteriorating, and argued that the costs of reducing ecological losses are usually higher than the benefits. Never mind that several of the world's foremost environmental scientists expressed more than mere scepticism towards Lomborg's grasp of their science: prominent publications such as *The Economist* promoted the book vigorously and wrote sermons on how scientists should practise their craft. People learning of my own work in developing ecological economics would ask, "And have you read Lomborg?" — implying, "Why have you thrown away so much of your working life?"

Things have changed over the past year. Former US vice-president Al Gore's film *An Inconvenient Truth* and the Fourth Report of the Intergovernmental Panel on Climate Change have given rise to great public concern, and many now regard global warming to be the central problem facing humanity. Lomborg's latest book, *Cool It*, is a response to that change in public perception. He doesn't question the science, which says that rising concentrations of greenhouse gases in Earth's atmosphere are affecting our climate system; he questions whether we should do much about it. If *The Skeptical Environmentalist* was the relentless prosecuting counsel, *Cool It* is the hard-headed but caring economist.

The book is a series of exercises in cost–benefit analysis, interspersed with quotes on climate change from the writings of famous people who should know better than to speak in hyperboles. Lomborg produces figures to show that it would be better to replace the Kyoto Protocol with strategies that encourage economic growth and blunt the harmful effects of climate change. Here is a sample: did you say Kyoto would result in fewer floods? Maybe, but it would reduce flood damage by only US$45 million a year, whereas building appropriate infrastructure could lower it by $60 billion a year. Didn't you also say that global warming would cause additional deaths from heatwaves? Yes, but what about the greater numbers who would not die of cold? Are you worried about deepening poverty in the tropics without Kyoto? You shouldn't be, because Kyoto would reduce the number of undernourished people in 2080 by only 2 million, whereas the United Nations proposes in its Millennium Development Goals to reduce the number by 229 million by 2015. What about more severe hurricanes? Well, Kyoto would reduce the increased annual damage by only 0.6%, whereas taking better precautions could lower it by 250%. And so on.

B. MCNEELY/GETTY IMAGES

Should we be spending more on protecting ourselves against the adverse effects of global warming?

Lomborg reports that Kyoto's annual cost would be $180 billion in foregone output, whereas the smart strategies he outlines, which would include an annual expenditure of $25 billion on research and development in clean technologies, would cost a mere $52 billion a year. By his reckoning, those strategies would limit the rise in concentration of carbon dioxide to 560 parts per million (p.p.m.) and the accompanying temperature rise to 4.7 °C. Smart strategies would cost far less than Kyoto, deliver higher economic growth worldwide, and markedly reduce poverty. From the vantage point of Kyoto, there is a free lunch to be had wherever you look.

You might say that the Kyoto Protocol was misconceived and that the world should develop a bolder programme of action, with much higher carbon taxes, international cooperation to reduce hunger, disease and habitat destruction, and development of clean technologies and ways to sequester carbon. But in Lomborg's view, doing more of a bad deal is rarely smart, so he doesn't countenance going beyond Kyoto. All this is spelt out in such a breezy, engaging style, it's hard not to find the arguments entirely reasonable.

Unfortunately, Lomborg's thesis is built on a deep misconception of Earth's system and of economics when applied to that system. The concentration of CO_2 in the atmosphere is now 380 p.p.m., a figure that ice cores in Antarctica have revealed to be in excess of the maximum reached during the past 600,000 years. If there is one truth about Earth we all should know, it's that the system is driven by interlocking, nonlinear processes running at different speeds. The transition to Lomborg's recommended concentration of 560 p.p.m. would involve crossing an unknown number of tipping points (or separatrices) in the global climate system. We have no data on the consequences if Earth were to cross those tipping points. They could be good, or they could be disastrous. Even if we did have data, they would probably be of little value because nature's processes are irreversible. One

implication of the Earth system's deep nonlinearities is that estimates of climatic parameters based on observations from the recent past are unreliable for making forecasts about the state of the world at CO_2 concentrations of 560 p.p.m. or higher. Moreover, the nonlinearities mean that doing more of a bad deal (Kyoto) may well be very good.

These truths seem to escape Lomborg. His cost–benefit analysis involves only point estimates of variables (interpreted variously as 'most likely', 'expected', and so forth), implying that he believes we shouldn't buy insurance against potentially enormous losses resulting from climate change. His concerns over the prevalence of malaria, undernutrition and HIV in today's world show that he is an egalitarian. There is, then, an internal contradiction in his value system, because if you are averse to inequality you should also be averse to uncertainty.

The integrated assessment models of Earth's system on which Lomborg builds his case are arbitrarily bounded on either side of his point estimates. It can be shown that if those bounds are removed (as they ought to be), even a small amount of uncertainty — when allied to only a moderate aversion to uncertainty — would imply that humanity should spend substantial amounts on insurance, even more than the 1–2% of world output that has been advocated. If the uncertainties are not small, standard cost–benefit analysis as applied to the economics of climate change becomes incoherent, even if those uncertainties are judged to be thin-tailed (gaussian, for example); this is because the analysis would say that no matter how much humanity chooses to invest in protecting Earth from passing through those later tipping points, we should invest still more.

Economics helps us to realize what we are able to say about matters that will reveal themselves only in the distant future. Simultaneously, it helps us to realize the limits of what we are able to say. That, too, is worth knowing, for limits on what we are able to say are not a reason for inaction. Lomborg's seemingly persuasive economic calculations are a case of muddled concreteness. ■

Partha Dasgupta is professor of economics at the University of Cambridge and fellow of St John's College, Cambridge. He is author of *Discounting Climate Change*, forthcoming in *Review of Environmental Economics and Policy*.

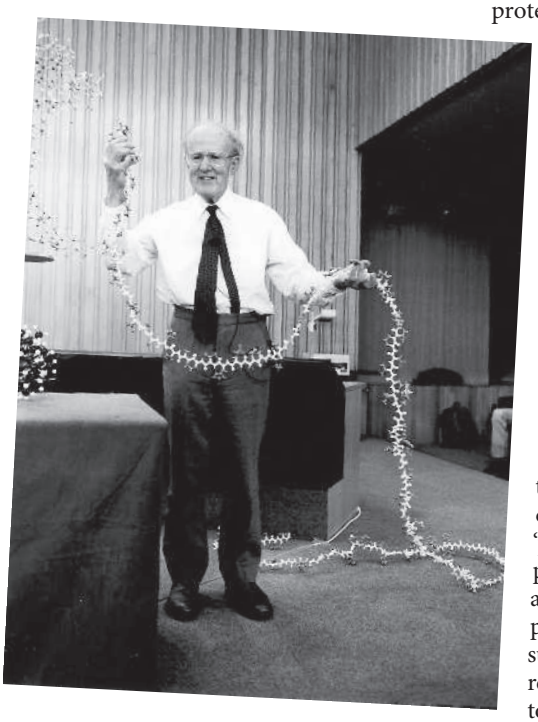
Max in three dimensions

Max Perutz and the Secret of Life
by Georgina Ferry
Chatto & Windus/Cold Spring Harbor Laboratory Press: 2007. 304 pp. £25/$39

Gregory A. Petsko

I have a problem with the title of this book, but it's almost my only quibble with this marvelous biography of one of the least known of the twentieth century's great scientists. By no measure could Max Perutz be said to have discovered the secret of life — a claim that might be defended for Gregor Mendel, or Charles Darwin and Alfred Russel Wallace, or James Watson and Francis Crick. The mechanism of oxygen's reversible binding to haemoglobin, which Perutz elucidated in 1970 after more than 25 years of work, doesn't even apply to most living organisms (and oxygen is deadly to most anaerobes). That said, Perutz did many extraordinary things, including winning the Nobel Prize in Chemistry in 1962 for solving haemoglobin's three-dimensional structure.

It is hard to write a biography of someone only recently deceased (Perutz died in 2002). The biographer must accurately portray someone who was known personally to many readers, yet at the same time expose previously hidden aspects of his or her character. Georgina Ferry achieved both in her biography of Dorothy Crowfoot Hodgkin, and now she has done it again in her engrossing account of the life and work of Max Perutz.



PERUTZ FAMILY

Perutz the showman: Max continued to captivate audiences well into his eighties.

Perutz, whom I knew well for 30 years, can't have been an ideal subject. No scandal colours his career; his personal life was stable and happy; his accomplishments were clear in terms of priority; he made a number of significant mistakes; his major discoveries are not easily understood by the layperson; and he lacked the forceful manner of a Crick or Bernal. But as the father of protein crystallography — arguably one of the greatest scientific advances of the last century — and the founder of the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, UK, his influence was enormous.

Ferry succeeds in bringing what could, in lesser hands, be considered a somewhat drab character sharply to life. As an Austrian Jew born during the First World War, Perutz left for Cambridge in 1936, where he joined the laboratory of one of the larger-than-life figures of modern science, John Desmond Bernal. Together with Dorothy Hodgkin, Bernal had, just two years earlier, taken the first X-ray diffraction photograph of a single crystal of a protein molecule, the digestive enzyme pepsin. This showed that, in principle, the extraordinary power of crystallography could reveal the atomic details of even large biological molecules. Undeterred by the scale of his task, Perutz ventured to use crystallography to unravel how haemoglobin could bind oxygen tightly enough to transport it around the bloodstream, yet release it when and where it was needed.

I doubt whether most people, even today, understand how pioneering this was. Determining the three-dimensional structure of proteins was a goal of Nobel-prize potential for several powerful research groups of that time, but none particularly cared what the protein was. Only Perutz had a greater aim. He wanted to understand the function of haemoglobin, which meant solving all the problems presented by this large, flexible protein. What would he have made of the recent International Structural Genomics Initiative, I wonder, which aims to turn out massive numbers of protein crystal structures without regard to biological or biochemical function?

Perutz's greatest achievement was demonstrating that the method of 'isomorphous replacement', previously used to solve the structures of small organic compounds, could be used to crack the 'phase problem' in protein crystallography. This is the problem of how to deduce a wave's phase component in diffraction patterns. This method made it possible to sum up the scattered X-ray waves in proper registration with each other and therefore to reconstruct the molecule's structure. It opened the way to solving the structure of any large crystalline molecule. Ferry explains the many false starts, and embarrassing errors, that led up to that moment, while allowing the reader to feel the frustrations and joys along the way. It's as good an account of a scientific breakthrough as you will find.

Haemoglobin's structure followed. More

than a decade later, Perutz showed that it was the minute movement of the iron atom into the plane of the haem group upon binding oxygen that triggered the shape change from the deoxygenated form of the protein. He once spent an hour explaining the new mechanism to me when I was a graduate student (even though he had just published it and must have talked about it many times before), as enthusiastically as a child with a new toy.

Perutz's sometimes child-like character is the surprise in Ferry's biography. Apparently, he had a desire to be praised for his discoveries, which sometimes manifested as petulance and led to ruthlessness towards competitors. But he was equally ready to confess his blunders, which endeared him to many younger scientists. Fragile health, combined with an acute sense of his failures, may have explained his reserve, particularly around the boisterous young molecular biologists at Cambridge. Ferry avoids the pop-psychology that permeates so many modern biographies, while offering insight into Perutz's temperament and behaviour.

For Ferry, one of Perutz's finest achievements was the creation in 1962 of the MRC Laboratory for Molecular Biology, which has produced an astounding number of Nobel laureates and Fellows of the Royal Society. She devotes an entire chapter to its history and to Perutz's unique, hands-off style of managing it. His skill in identifying and nurturing talent at a time when molecular biology was just starting out is one of the things that makes Perutz a central figure in modern scientific history.

Ferry doesn't end the book with Perutz's death from Merkel cell carcinoma, just days after submitting his manuscripts on the structure of the protein aggregates in Huntington's disease. Instead, she follows it with a chapter about his avocation as a writer of popular essays about science and society. This is a masterstroke, because his wise and witty writings present Perutz to us at his most candid, and so the chapter sums up the book, and the man, very nicely.

Ferry has mined gold in the lives of two of the founders of structural biology; I can't wait to see whom she tackles next. Frederick Sanger, one of only four people to win two Nobel prizes? Or how about William H. Bragg or Max von Laue? ■

Gregory A. Petsko is professor of biochemistry and chemistry at Brandeis University, 415 South Street, Waltham, Massachusetts 02454-9110, USA, and adjunct professor in the Department of Neurology and Center for Neurologic Diseases, Brigham & Women's Hospital, Harvard Medical School.

One man and his molecule

Piccole Visioni: La Grande Storia di una Molecola
by Marta Paterlini
Codice Edizioni: 2006. 263 pp. €19

Ermanno Gherardi

Max Perutz and his 1959 model of oxygenated haemoglobin is one of the iconic images of twentieth-century biology. It encapsulates a journey that began in 1936 when, armed with a degree in chemistry from the University of Vienna, Perutz moved to Cambridge to work as John Desmond Bernal's research student on the task of solving protein structures at atomic resolution using X-ray crystallography. At the time, Bernal and his former research student, Dorothy Crowfoot Hodgkin, were probably the only two people to believe that the atomic structure of a protein was within reach — having themselves obtained promising diffraction patterns from hydrated crystals of pepsin just two years earlier. Perutz was thus charged with the responsibility of realizing Bernal's dream. A year later, he opted to work on haemoglobin, the oxygen-transporting protein in red blood cells, a study that lasted for the next 60 years.

Piccole Visioni ('Small Visions'), written in Italian by Marta Paterlini, narrates the story of how Perutz arrived at the atomic structure of haemoglobin and, from there, at the finely tuned mechanism that regulates oxygen binding, transport and off-loading at its destination. The book also provides a vivid account of Perutz's life and his role in founding the Medical Research Council (MRC) Laboratory of Molecular Biology, the main institution responsible for the birth of 'new biology' in the second half of the twentieth century.

Perutz's early years in Cambridge saw him concentrating on the crystallographic analysis of haemoglobin. He was under considerable personal strain at the time, because his parents had lost their home and property following Hitler's invasion of Austria; they arrived as refugees in Cambridge in 1939. Perutz himself, a potential 'enemy alien', was initially interned and deported to Canada until January 1941, when — in a dramatic change of heart — the British government involved him in a secret war project to produce reinforced ice to act as floating airfields (but which were never built).

SVENSKT PRESSFOTO/PHOTOSHOT

A glowing Perutz at the 1962 Nobel ball, with his wife Gisela.

Throughout the war years, funding from the Rockefeller Foundation in the United States and later from Imperial Chemical Industries in the United Kingdom — obtained with the help of William Lawrence Bragg, Cavendish professor from 1938 — enabled Perutz to carry on with his work. In 1947, Bragg persuaded the MRC to set up a "unit for the study of molecular structure of biological systems" involving himself, Perutz, John Kendrew and two assistants. This unit later became the present MRC Laboratory of Molecular Biology in Cambridge.

The early chapters of *Piccole Visioni* cover these events and introduce the basic concepts of crystallography and the problems that were faced in analysing protein crystals at the time. The rest of the book is taken up with Perutz's discovery of the basic features of protein structure through his studies on haemoglobin. A breakthrough came in 1951, when he experimentally confirmed the structure of the α-helix shortly after Linus Pauling had proposed its existence on theoretical grounds. Two years later, Perutz developed the technique of isomorphous replacement for protein crystallography (see 'Max in three dimensions', opposite).

Perutz's method enabled him to determine a low-resolution structure of haemoglobin and helped Kendrew to solve the structure

of myoglobin in the late 1950s, for which they shared the Nobel Prize in Chemistry in 1962. *Piccole Visioni* extends these developments with an account of the subsequent higher-resolution structures of deoxygenated and oxygenated haemoglobin.

Piccole Visioni offers a lively and penetrating insight into the life and work of Perutz. Paterlini writes clearly and her book is well researched, successfully portraying both Perutz's science and his humanity, honesty and ability to reach a decision by persuasion. It has been argued that these qualities were instrumental in developing the MRC Laboratory of Molecular Biology into a highly original and successful research institution — indeed, Perutz's style of management survived his retirement as chairman of the laboratory in 1979. Perutz also played an early and important part in the genesis of the European Molecular Biology Organisation (EMBO).

Piccole Visioni does not deal with another of Perutz's legacies — namely, his writings for the public, now a commonplace initiative. He wrote several beautiful short essays collected in books such as *Is Science Necessary?* (1989), *Science is Not a Quiet Life* (1997) and *I Wish I'd Made You Angry Earlier* (1998, 2002).

Piccole Visioni is among the first of several books that address Max Perutz's gigantic contribution to twentieth-century biology and science culture. Paterlini sets the scene for these and her book will hold its place among them. I hope that it will soon become available in other languages. ■

Ermanno Gherardi is at the Medical Research Council Centre, Cambridge CB2 2QH, UK.

SCIENCE IN CULTURE

Heaven in grains of sand

Nanoscientists and Tibetan monks unite to explore the mysteries of the mandala.

Martin Kemp

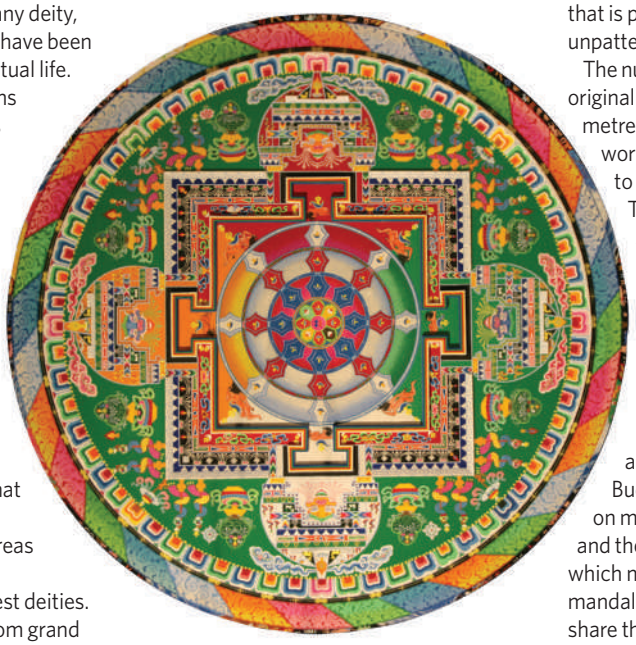
Western religious art from the time of ancient Greece has generally relied on the representation of the human figure. But in many world cultures, especially those that proscribe the literal depiction of any deity, symbolic schemata and patterns have been used to express the truths of spiritual life.

Over the centuries, such patterns have evolved extraordinary levels of intricacy. Islamic tiling, famed for its remarkable mathematical symmetries, provides the most commonly cited example of age-old pattern generation that continues to engage our intellectual and aesthetic interests.

Less familiar, but no less striking, are the many Buddhist variations of the mandala, which symbolizes the Universe. The outer region of the mandala — 'that which encircles' — expresses the cosmological world system, whereas the inner zones characteristically progress to the realm of the highest deities. Dense, with map-like features, from grand structures to tiny symbolic details, mandalas serve the spiritual exercise of diligently sustained contemplation.

The most astonishing are the Tibetan mandalas composed from coloured grains of loose sand (www.imagesandmusic.nl/dhtml/Mandala.htm#). The time and patience required to create them is incredible to anyone involved in the frenzy of modern life. The meticulous, laborious act of making it is itself a key part of the discipline of timeless meditation. It is, as William Blake wrote, "to see heaven in a grain of sand".

This compound of extreme physical minuteness and grand cosmological structure has recently provided the rationale for the presence of a group of Tibetan monks in a leading nanotechnology laboratory. The initiative is led jointly by an artist and a chemist at the University of California in Los Angeles — Victoria Vesna from the Department of Design | Media Arts, and nanoscientist James Gimzewski from the Department of Chemistry and Biochemistry (see http://nano.arts.ucla.edu/mandala).



V. VESNA

The Chakrasamvara Mandala is made of coloured sand. Its intricacies have been revealed down to the molecular level.

They have collaborated with monks from the Ghaden Lhopa Khangsten Monastery in India to explore a Chakrasamvara mandala. The mandala, called the 'wheel of great bliss', encircles the palatial residence of the deity Heruka Chakrasamvara. It places particular emphasis on the female ideal of wisdom.

As an extension of the monks' rendering of the cosmological whole from the tiniest grains, Gimzewski has used optical and scanning electron microscopy to delve into progressively smaller features of the sand mandala, right down to the molecular level. Microscopic images across this range were then blended with a sequence of zoomed photographs to produce a continuous visual journey from the whole mandala into ever-finer details of its physical composition. The result is a seamless 15-minute sequence that is projected onto a circular bed of flat, unpatterned sand.

The numbers involved are awesome. The original sand mandala was two-and-a-half metres in diameter and took four monks working for eight hours a day four weeks to complete in Gimzewski's laboratory. The final computer output comprised 30,000 individual frames containing 900 gigabytes of data. Thirty-six computers were pressed into service to render the images over the course of two days, and nine computers completed the recomposition of the continuous sequence.

There is something very beautiful and moving in this holy alliance of Buddhist spiritual patience, founded on minute care and untiring repetition, and the unholy processes of iteration of which modern computers are capable. The mandala-makers and the nanoscientists share the wonder of scale, involving countless parts to compose the ordered whole. We can sense the way in which religious contemplation of a time-honoured kind and modern technological science are, in their different ways, reaching out to the edges of infinity.

This is the aesthetic realm of the sublime. It is inhabited by all those who stand in awe at the wonder of the Universe and in thrall to the varied mental capacities we use to make sense of what we see and feel. ■

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK.

The piece is to be exhibited at the Maison Européenne de la Photographie in Paris until 30 September then at the Singapore Science Centre until 8 December 2011.

NEWS & VIEWS

EXTRASOLAR PLANETS

The one that got away

Jonathan Fortney

Hanging around a star that has passed through its red-giant phase doesn't seem a likely place for a planet. But one planet apparently managed to avoid being engulfed by its bloated star — might others, too?

One of the most surprising things that we have learnt about planets during the past 15 years is that they can turn up almost anywhere. On page 189 of this issue, Silvotti *et al.*[1] expand our planet-finding horizon still farther: they have discovered a massive gas-giant planet orbiting an old star in a rare, late stage of the star's evolution. This system allows us to start examining what will happen to planets around stars such as our own Sun as they too evolve and grow old.

The parent star of the newly discovered planet is known as V 391 Pegasi. During its time as a middle-aged star, it had a mass similar to that of the Sun, and stably fused hydrogen into helium for billions of years. But once its core had burnt fully to helium, the star entered its red-giant phase, expanding in radius by more than 100 times. In 5 billion years' time, our Sun will suffer much the same fate, expanding out perhaps as far as to Earth's orbital distance.

Unusually, however, and for reasons that are not altogether clear, V 391 Pegasi lost its hydrogen-rich surrounding envelope in a strong, outflowing wind at the end of its red-giant phase. It was left with only a thin skin of atmosphere atop its burnt-out helium core. This type of compact, dense star is known as a B-type subdwarf, has a surface temperature of around 30,000 kelvin, and is powered not by hydrogen, but by the fusion of helium into carbon.

V 391 Pegasi is one of 40 stars in its rare class known to pulsate. The pulsations, which have periods of several minutes[2], give us clues to the structure of these stars[3] (similarly, the characteristic frequencies of the Sun's pulsations have told us much about its interior). But the pulsations also make it unusually easy to detect planets. The trick is to make precise observations of the timing of the pulses: the gravitational pull of the planet orbiting the star leads to subtle shifts in the distance of the star from Earth, which is reflected in the pulses' arrival times here.

This method has yielded the most exotic planetary systems discovered so far. The first planets found outside our Solar System ('exoplanets' in the jargon), which are in orbit around an ultra-dense neutron star, were discovered in this way[4]. They are in orbit around a pulsar — a radio-wave-emitting, ultra-dense neutron star that is a remnant of a supernova. In V 391 Pegasi's case, as Silvotti *et al.*[1] report, the clincher for the planetary hypothesis is two distinct pulsation modes, each with frequencies around 350 seconds, that vary in phase on the same 3.2-year timescale — probably the time the planet takes to orbit its star (Fig. 1).

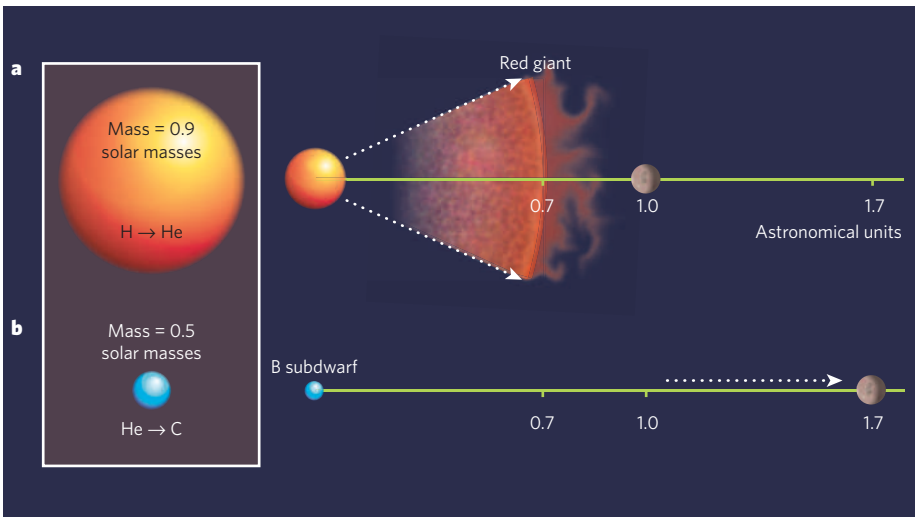


Figure 1 | A picture of things to come. a, When the star V 391 Pegasi was middle-aged — as our Sun is now — it fused hydrogen into helium, and its planet, identified by Silvotti *et al.*[1], sat at 1 astronomical unit, the same distance Earth is from the Sun. As the star evolved into a red giant, it expanded more than 100 times in radius to 70% of the star–planet distance. **b**, But the star later lost almost half of its mass, becoming a B-type subdwarf, for unknown reasons; it is possible that the presence of the planet itself played a part. With the star's gravitational pull reduced, the planet migrated to a more distant orbit, where it now resides, causing a variation in the phase of pulsed signals from V 391 Pegasi with a period of 3.2 years.

The discovery of such an unusual star–planet system represents an opportunity to understand both the star and the planet better. Astronomy is not an experimental science — we can't make a solar system to test a hypothesis — and so astronomers are left to search the skies to find a statistically significant number of samples, at various evolutionary stages, to construct consistent physical interpretations. At present, Silvotti and colleagues' discovery is the only planetary system known to have survived past its parent star's red-giant phase.

The formation of B subdwarfs is poorly understood: 98% of stars that reach the red-giant phase do not undergo the catastrophic mass loss characteristic of V 391 Pegasi and its brethren[5]. The planet's effect on the star's fate is unclear, and could be minimal. But one suggestion[6] is that a massive orbiting companion might deposit angular momentum and energy onto a star's hydrogen envelope, and so enhance its mass loss. To demonstrate convincingly that this mechanism is the dominant mode for forming these stars, however, many more B subdwarfs would need to be found harbouring a planet.

And predicting where planets can and cannot be is a tricky business. For instance, since 1995 we have learnt that in their Sun-like middle age about 1 in 100 'normal' stars have gas-giant planets of a similar mass to Jupiter orbiting within only 5–10 radii of their surface[7]. With a sample so far of almost 250 (and counting) planets around such stars, we are slowly starting to understand the diverse architectures of extrasolar planetary systems. What the configuration of planetary systems

might be like around older stars further along in their evolution has been analysed in several studies. These have focused primarily on how planetary orbits might move outwards as stars lose mass, or inwards as planets are dragged in and consumed in the outer envelope of a bloated old star[8,9], and, more recently, on planetary evaporation in the face of intense stellar irradiation[10].

As in all of astronomy, further progress will involve finding additional objects so that we can understand planets around evolved stars as a class. If searches around evolved pulsating stars such as B subdwarfs and around still older, more-compact white dwarfs yield more planets, astronomers will be on the way towards understanding how stellar evolution affects the architecture of planetary systems. This will shed light not only on our own Solar System, in which Mercury, Venus and perhaps Earth will eventually be engulfed by the red-giant Sun[9], but also on the diverse array of planetary systems that are our Galactic neighbours. ■

Jonathan Fortney is at the NASA Ames Research Center, MS 245-3, Moffett Field, California 94035, USA.
e-mail: jfortney@arc.nasa.gov

1. Silvotti, R. *et al. Nature* **449,** 189–191 (2007).
2. Kilkenny, D. *Commun. Asteroseismol.* **150,** 234–240 (2007).
3. Charpinet, S. *et al. Astron. Astrophys.* **459,** 565–576 (2006).
4. Wolszczan, A. & Frail, D. A. *Nature* **355,** 145–147 (1992).
5. Han, Z., Podsiadlowski, P., Maxted, P. L. F., Marsh, T. R. & Ivanova, N. *Mon. Not. R. Astron. Soc.* **336,** 449–466 (2002).
6. Soker, N. *Astron. J.* **116,** 1308–1313 (1998).
7. Butler, R. P. *et al. Astrophys. J.* **646,** 505–522 (2006).
8. Rybicki, K. R. & Denis, C. *Icarus* **151,** 130–137 (2001).
9. Duncan, M. J. & Lissauer, J. J. *Icarus* **134,** 303–310 (1998).
10. Villaver, E. & Livio, M. *Astrophys. J.* **661,** 1192–1201 (2007).

EPIGENETICS

Perceptive enzymes

Anne C. Ferguson-Smith and John M. Greally

Adding methyl groups to DNA is a way of regulating some genes and genomic sequences. Structural analysis reveals that the enzyme complex that mediates this process shows unexpected sequence specificity.

Imprinted genes are a small but developmentally important set of genes whose expression depends on the parent from which they are inherited. So, for some of these genes only the maternally inherited copy is expressed, and for others only the copy inherited from the father is expressed. Such selective gene expression is regulated by selective addition of methyl groups to the two equivalent parental chromosomes during the development of gametes (eggs and sperm); these chemical marks are then propagated in the resulting offspring[1]. In vertebrates, methylation of DNA, which is usually associated with the shut-down of local gene expression, is mediated by DNA methyltransferase enzymes. A question that has puzzled researchers is how these enzymes discriminate between different sequences within the genome. On page 248 of this issue, Jia *et al.*[2] provide some clues, reporting that DNA methyltransferases show remarkable sequence specificity.

In higher organisms, DNA methyltransferases mainly methylate cytosine–guanine (CG) dinucleotides. But regions in the genome with the greatest density of potential CG dinucleotide targets, known as CpG islands, are generally unmethylated in normal cells. And, to add to the paradox, parasitic genomic sequences called transposable elements, which are not particularly rich in CG dinucleotides[3], are usually methylated.

To understand how DNA methyltransferases interact with their target DNA, Jia *et al.*[2] set out to resolve the structure of one DNA methyltransferase, Dnmt3a, in a complex with its regulatory factor Dnmt3L. For successful DNA methylation, Dnmt3a and Dnmt3L must work together, and previous studies[4–6] had shown that if either of the genes encoding Dnmt3a or Dnmt3L is mutated, DNA sequences that regulate genomic imprinting fail to establish proper methylation.

Jia and colleagues find[2] that Dnmt3a and Dnmt3L form a tetramer in the order Dnmt3L–Dnmt3a–Dnmt3a–Dnmt3L; they describe the structure as resembling the wings of a butterfly. More noteworthy, however, are the authors' findings that the central Dnmt3a dimer in the tetramer can methylate two CG dinucleotides in one binding event, and that it preferentially methylates pairs of CGs that are 8–10 base pairs apart, with other CGs in between remaining unmethylated. Earlier work[7] had also shown that, under these conditions, the enzyme is non-processive, meaning that it works where it lands rather than moving along a DNA sequence. This pattern of preferred methylation suggests the potential for DNA-sequence specificity. The question is whether this previously unsuspected specificity contributes to the methylation of genomic imprints.

The authors explored this intriguing possibility, looking for CG periodicity in imprinting control regions that contain differentially methylated sequences. They found that in these regions CG dinucleotides do indeed occur with a periodicity of 8–10 base pairs. By contrast, a sample of ten unmethylated CpG islands from chromosome 21 did not have the same CG periodicity. Furthermore, the authors detected this periodicity only in the 12 regions targeted for methylation during egg development, and not in the 3 regions acquiring methylation during sperm formation. These observations, which attest to a remarkable specificity of the Dnmt3a–Dnmt3L

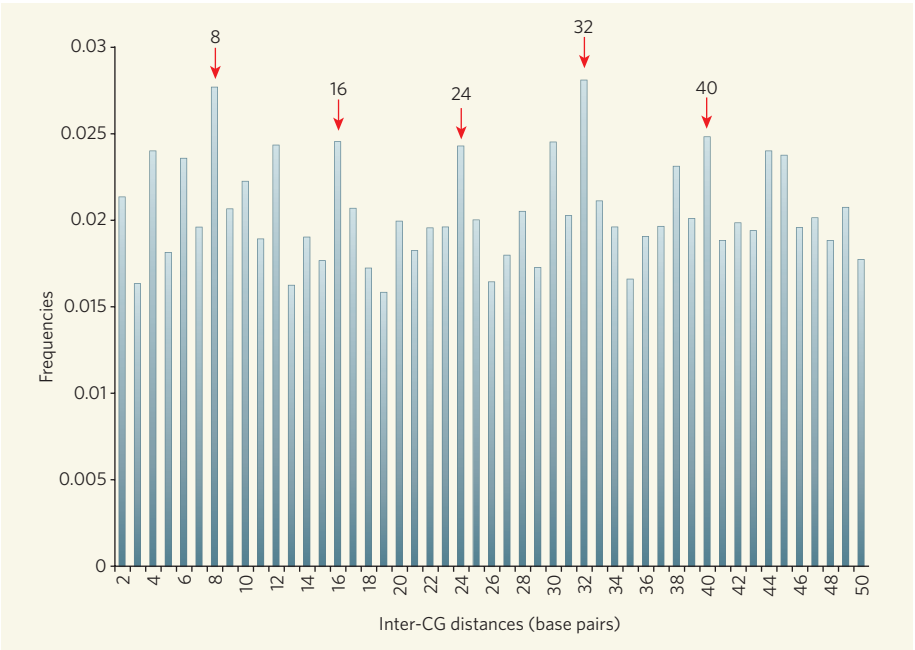


Figure 1 | Covert signals. To complement the findings of Jia *et al.*[2], we used the University of California, Santa Cruz genome browser[9] to search the human genome for the periodicity of CG dinucleotides. Our results indicate that a CG periodicity of eight base pairs — and, therefore, potential target sites for methylation by the Dnmt3a–Dnmt3L enzyme complex — is very common throughout the human genome.

complex, fit nicely with earlier observations[4–6] showing that mutations in the gene encoding Dnmt3L affect DNA methylation mainly during egg development.

Of the several questions that Jia and colleagues' work raises, one is the frequency with which a CG periodicity of 8–10 base pairs occurs across the genome. The authors did not detect such periodicity in ten randomly selected CpG islands they tested. However, using the same approach that they used, we searched the entire human genome and found a striking over-representation of CG dinucleotides with an eight-base-pair periodicity (Fig. 1). The potential targets for the Dnmt3a–Dnmt3L complex therefore occur throughout the genome and do not seem to be limited to differentially methylated regions containing imprinted genes. Consequently, this enzyme complex might have a broader potential role than methylating only the restricted set of genes undergoing genomic imprinting.

A recent study[8] has shown that Dnmt3L is prevented from interacting with DNA when it is associated with a methylated form of the core histone, H3 (DNA wraps around such histone proteins to form chromatin). This potential protection from DNA methylation on the basis of a particular histone modification suggests another layer of genome recognition, this time linking the DNA methylation machinery to the modification state of histones.

By integrating structural biology with DNA analysis, Jia and colleagues[2] have cracked another component of the complex epigenetic system of modifications that modulate the function of the genome without affecting its sequence. They have revealed an underlying functional periodicity to the seemingly random distribution of CG dinucleotides within crucial regulatory sequences and illuminated how Dnmt enzymes can perceive these patterns and use them to target their activity preferentially. This raises the possibility that different patterns of CG periodicity might reveal different functional specificities within the genome. ■

Anne C. Ferguson-Smith is in the Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. John M. Greally is in the Departments of Medicine (Hematology) and Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York 10461, USA.
e-mails: afsmith@mole.bio.cam.ac.uk; jgreally@aecom.yu.edu

1. Edwards, C. A. & Ferguson-Smith, A. C. *Curr. Opin. Cell Biol.* **19,** 281–289 (2007).
2. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. & Cheng, X. *Nature* **449,** 248–251 (2007).
3. Fazzari, M. J. & Greally, J. M. *Nature Rev. Genet.* **5,** 446–455 (2004).
4. Bourc'his, D., Xu, G.-L., Lin, C.-S., Bollman, B. & Bestor, T. H. *Science* **294,** 2536–2539 (2001).
5. Bourc'his, D. & Bestor, T. H. *Nature* **431,** 96–99 (2004).
6. Kaneda, M. *et al. Nature* **429,** 900–903 (2004).
7. Gowher, H. & Jeltsch, A. *J. Mol. Biol.* **309,** 1201–1208 (2001).
8. Ooi, S. K. T. *et al. Nature* **448,** 714–717 (2007).
9. Kent, W. J. *et al. Genome Res.* **12,** 996–1006 (2002).

CHEMISTRY

Molecular socks in a drawer

Michael D. Ward

Dynamic combinatorial chemistry is a deft way to identify the most stable forms in a complex mixture of interconverting compounds. Even more cunningly, it can also be used to sort related molecules by crystallization.

As chemists seek to produce increasingly complex molecules and molecular assemblies, it has become evident that complexity often comes at a price — the many different reaction pathways can generate unwanted mixtures and side products. This is particularly true when the reaction pathways and products are similar energetically.

One way around this problem is 'dynamic combinatorial chemistry', which involves looking for new molecules using 'libraries' of constituents in solution. These molecular constituents can change into one another through one or more reversible chemical processes, ranging from covalent-bond formation to hydrogen-bonding to the formation of metal complexes[1]. Writing in the *Journal of the American Chemical Society*, Hutin *et al.*[2] report a new twist, describing a complex library of chiral constituents that 'self-sort' through their differing solubilities. The result is a single, pure product that crystallizes out of solution under thermodynamic control.

Chiral molecules can exist in two different

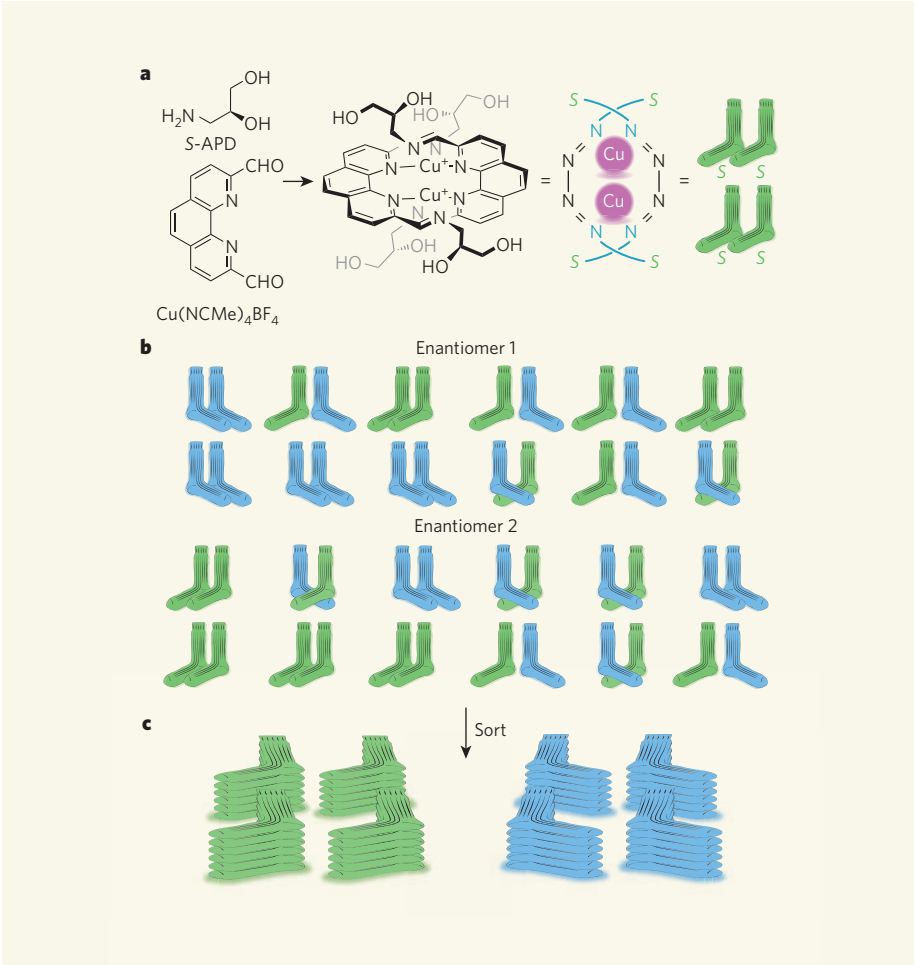


Figure 1 | Sorted. a, Reacting a chiral form of 1-amino-2,3-propanediol (here the left-handed version, *S*-APD) and a phenanthroline ligand with copper ions in solution produces a complex molecule with four chiral sites — regard them as four socks facing in the same direction. **b**, As Hutin *et al.* show[2], it gets more interesting when, rather like putting an assortment of socks in a drawer, a mixture of the right-handed (*R*-APD) and left-handed (*S*-APD) versions of the molecule is used. The result is a complex library of six chiral forms and their respective enantiomers (here, green, left-facing socks equate to *S*-APD and blue, right-oriented socks to *R*-APD). **c**, Startlingly, when the mixture is left to stand, racemic crystals form that contain two neatly sorted piles of two distinct molecular isomers — one with four *R*-APD sites and the other with four *S*-APD sites.

conformations — right- and left-handed mirror images of each other, known as enantiomers. Such molecules crystallize primarily in one of two forms: as a 'racemate', meaning that the crystal is an ordered arrangement of equal numbers of both enantiomers; or as two sets of distinct enantiomorphic crystals, each containing only left-handed molecules or only right-handed molecules. The most famous example of this second process is Louis Pasteur's seminal discovery of the spontaneous separation of sodium ammonium tartrate into left- and right-handed crystals[3].

In the laboratory, racemic crystals occur more frequently than enantiomorphic crystals. This has been attributed to Wallach's rule[4,5], which states that racemic crystals tend to be denser — the left and right hands are arranged around centres of symmetry, and their packing is thus more efficient — than their enantiomorphic counterparts, and therefore more stable. Furthermore, symmetry constraints limit the number of favourable packing arrangements available to chiral molecules.

Hutin *et al.*[2] used the low solubility of a crystal racemate to retrieve the left-handed and right-handed versions of a specific pair of chiral molecules from a mixture of many closely related, interconverting conformations of the same molecule, which are likely to have very similar energies. The authors started by attaching a chiral building-block (1-amino-2,3-propanediol, or APD) to opposite ends of each of two bridging phenanthroline (phen) ligands. These ligands are themselves bound together by two copper ions (Fig. 1a). The resulting 'dicopper double dihelicate' molecule has four APD chiral centres, each of which can be either left handed (written *S*) or right handed (*R*).

Things became interesting when, rather like putting a bundle of assorted socks in a drawer, the authors used a racemic mixture of *S*-APD and *R*-APD building-blocks. Under this condition, nuclear magnetic resonance studies indicated the presence of a mixture of numerous isomers. Thus (using a hyphen to denote the phen bridge between two chiral centres and a colon to separate the two different phen ligands), all the forms *R*-*R*:*R*-*R*, *S*-*R*:*R*-*R*, *S*-*S*:*R*-*R*, *S*-*R*:*R*-*S*, *S*-*R*:*S*-*R* and *S*-*S*:*R*-*S* were present in the solution, together with their partners of opposite chirality, *S*-*S*:*S*-*S*, *R*-*S*:*S*-*S*, *R*-*R*:*S*-*S*, *R*-*S*:*S*-*R*, *R*-*S*:*R*-*S* and *R*-*R*:*S*-*R* (Fig. 1b).

After being left to stand for two weeks, this racemic solution produced crystals. Even when large quantities of crystalline material were retrieved, the product was always a racemate containing only the enantiomeric *R*-*R*:*R*-*R* and *S*-*S*:*S*-*S* forms, in equivalent amounts and arranged in separate columns within the crystal (Fig. 1c). This surprising finding seems to indicate that these two forms are removed continuously and exclusively by crystallization, to be continuously replenished in the solution through the conversion of the other forms listed above. This in turn requires reversible exchange of *S*-APD and *R*-APD building-blocks among the molecules. The overall effect is a self-sorting process, in which a single crystalline racemate — presumably the most stable, least soluble version — forms from many possibilities.

This is a remarkable result, as it illustrates that crystallization can produce a single outcome among many possibilities from a mixture under thermodynamic control, even when the energetic differences between the many possible single outcomes is probably small.

A particularly attractive feature of a dynamic combinatorial library is the ability to adjust its composition, and so the stability ranking of its components, through changes in external factors such as temperature, pressure and light exposure. Libraries containing large numbers of interconverting chiral components, such as that described by Hutin *et al.*[2], represent a unique opportunity to explore the factors that determine whether a racemate or its corresponding enantiomorphs will be formed[6–9], a poorly understood phenomenon. Such libraries might also prove useful for optimizing and regulating crystal polymorphism and crystallization outcomes in general[10], a feature that would interest academic and commercial laboratories alike — particularly those dealing with pharmaceutical compounds and other specialist chemicals. ■

Michael D. Ward is at the Molecular Design Institute, Department of Chemistry, New York University, 100 Washington Square East, New York, New York 10003-6688, USA.
e-mail: mdw3@nyu.edu

1. Corbett, P. T. *et al. Chem. Rev.* **106,** 3652–3711 (2006).
2. Hutin, M. *et al. J. Am. Chem. Soc.* **129,** 8774–8780 (2007).
3. Pasteur, L. *Ann. Chim. Phys.* **24,** 442–459 (1848).
4. Wallach, O. *Justus Liebigs Ann. Chem.* **286,** 90–143 (1895).
5. Brock, C. P., Schweizer, W. B. & Dunitz, J. D. *J. Am. Chem. Soc.* **113,** 9811–9820 (1991).
6. Schipper, P. E. & Harrowell, P. R. *J. Am. Chem. Soc.* **105,** 723–730 (1983).
7. Custelcean, R. & Ward, M. D. *Cryst. Growth. Des.* **5,** 2277–2287 (2005).
8. Coquerel, G. *Top. Curr. Chem.* **269,** 1–51 (2006).
9. Jacques, J., Collet, A. & Willen, S. H. *Enantiomers, Racemates, and Resolutions* (Krieger, Malabar, FL, 1994).
10. Bernstein, J. *Polymorphism in Molecular Crystals* (Oxford Univ. Press, 2002).

ECOLOGY

Scaling laws in the drier

Ricard Solé

The vegetation of arid ecosystems displays scale-free, self-organized spatial patterns. Monitoring of such patterns could provide warning signals of the occurrence of sudden shifts towards desert conditions.

Once upon a time the Sahara was green — it was covered by vegetation. The evidence for this comes from many different sources, including the former existence of lakes. Around 5,500 years ago, the wet environmental conditions suddenly came to an end. Despite the absence of abrupt, external climatic change, plant productivity declined and the topsoil was lost. Eventually, the green Sahara became the desert Sahara that we know today[1]. The changes experienced by the biosphere over the past century, particularly increased desertification due to rising temperatures and declining rainfall, have raised concerns about the possibility of rapid shifts from green to desert states[2,3]. Arid and semi-arid ecosystems cover one-third of Earth's land surface, so there is a pressing need for quantitative ways to help forecast such shifts.

In this issue, Scanlon *et al.*[4] (page 209) and Kéfi *et al.*[5] (page 213) explore the problem of how vegetation in semi-arid ecosystems is organized in space and time. These studies point the way to how forecasting might be achieved. They involve analyses of the size distribution of vegetated patches in the Kalahari Desert[4] and in three different areas of the Mediterranean basin[5], and they cover different spatial scales and types of vegetation.

A notable finding by both groups is that the size distribution of vegetation clusters in undisturbed plots falls off as a power law: most patches of vegetation are of small size, but a few of them are very large. Specifically, if S is the size of a given vegetation cluster, then its frequency decays as $1/S^{\gamma}$ (with scaling exponents within the range $1<\gamma<2$). Such power laws occur in other types of ecosystem[6] and are a fingerprint of self-organization: that is, they are the result of internal dynamic processes driven by local interactions. This principle applies to the field data reported by both Scanlon *et al.*[4] and Kéfi *et al.*[5]. It indicates that plant interactions play a central role in shaping these ecosystems, which as a whole are characterized by productivity levels that largely depend on precipitation.

The authors also identify the origin of the mechanism underlying self-organization: a process of 'local facilitation' among plants, set against the background of overall control by water availability. Water is the limiting resource, but short-range interactions among plants involve positive effects that are a necessary condition for power laws to exist. The plants create a local environment that minimizes water run-off and facilitates the survival of other plants and seeds (Fig. 1, overleaf).

NATURE|Vol 449|13 September 2007

DIGITAL VISION/GETTY

Figure 1 | Mutual benefits. Semi-arid ecosystems, such as the Kalahari (pictured here), are characterized by harsh conditions dominated by water availability. In a process of positive feedback, called 'local facilitation'[4,5], plants that are specialized for such conditions create microenvironments that help other plants to survive. In consequence, neighbouring bare, degraded ground (D in Fig. 2a) can revert to fertile ground (E in Fig. 2a).

The two groups[4,5] support this claim by means of modelling with cellular automata, which in both cases successfully reproduces the observed spatial patterns and their scaling-law behaviour. These computer simulations involve the construction of a regular grid of cells, each of which has a particular state. In this case, the basic model considers three possible states: vegetated (V), empty (E) and degraded (D) (Fig. 2a). The first two designate fertile patches that are or are not occupied by plants. The third refers to degraded soil that cannot be colonized by plants. These three types of patch are related to each other through transitions governed by dynamical rules. Some transitions are affected by the presence of neighbouring vegetated patches. Here the positive effect of local facilitation is clear: it allows bare desert patches to revert to fertile soil that can later be colonized by local seedlings. This type of model leads to scaling laws under a wide range of conditions. Importantly, the work provides the first well-documented example of so-called robust criticality theory[7].

An implication of these results is that it might be possible to predict the transition from a vegetated to a desert state. It has been conjectured that arid ecosystems might suddenly shift towards a desert condition as external conditions deteriorate[2,3]. The typical example of a changing condition cited is decreasing rainfall. But increased grazing pressure by animals has a similar impact, and Kéfi *et al.*[5] reveal that more intense grazing leads to a departure from the power-law behaviour, with large patches becoming less and less common. Models consistently predict that such changes in size distribution might be warning signals of the approach of a transition to the desert state (Fig. 2b). The seriousness of such a possibility is highlighted by the fact that some of these transitions are catastrophic and largely irreversible[2].

The findings presented in these two papers[4,5] are compelling and seem robust. They open up the prospect of testing previous models against reality, and of forecasting future changes in arid ecosystems. For those purposes, however, we will need a rigorous theory that can help in explaining the exact origins of the scaling behaviour and in identifying other spatial measures that are needed to properly test the accuracy of a model. Previous work has been couched in terms of average quantities[7], and we require a modelling approach that allows predicted scaling exponents (and their variability) to be estimated more precisely and interpreted in terms of measurable parameters (such as water availability or degree of facilitation).

Moreover, changes in plant productivity are only the first layer in understanding how the whole food web will react to external stresses, and theoretical work will also need to take account of other levels in the food web. In this context, organisms inhabiting arid regions tend

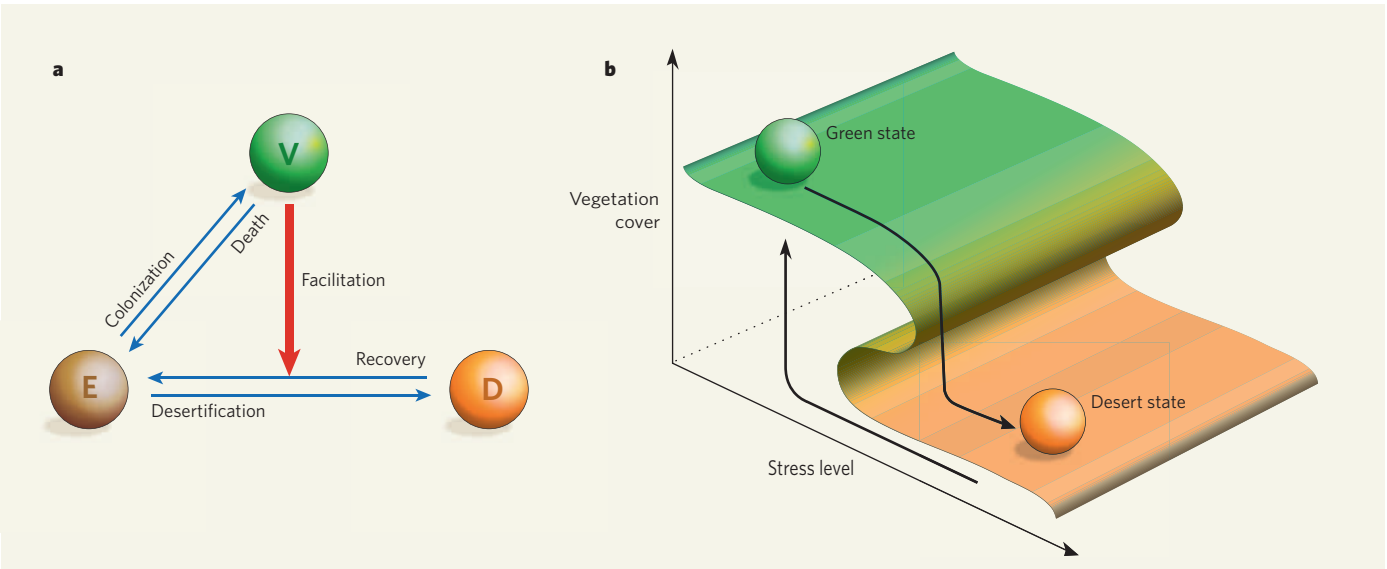


Figure 2 | Vegetation states in arid and semi-arid ecosystems. a, Basic interactions between different states of ecosystem patches — vegetated (V), empty (E) and degraded (D). Transitions from one state to another are possible at given rates (blue arrows), some of which are enhanced by the presence of neighbouring vegetation and the resulting improvement in physical conditions through the process of local facilitation (red arrow). **b**, A consequence of the nonlinear interactions between the three types of patch is the possibility of bistable behaviour between the well-vegetated (green) state and the desert state. Theoretical models show that a sudden transition from the first state to the second can occur under a continuous change in external stress, such as decreased water or increased grazing[5]. Like a marble rolling over a folded surface, the ecosystem changes continuously until a critical boundary is reached. At that point, the system suddenly shifts to the new state.

to have high levels of genetic diversity within species, and to be endemic to their particular ecosystem. They are thus of great relevance in considering biodiversity. A fuller theory of scaling behaviour is required to provide firmer connections between predicted and observed patterns of desertification — and so to provide a better understanding of the nature of transitions between green and desert phases. These belong to the family of non-equilibrium phase transitions seen in several different fields of research[8,9]. There is a great opportunity here for interdisciplinary work that would have potentially far-reaching consequences in conservation biology. ■

Ricard Solé is in the ICREA-Complex Systems Laboratory, Universitat Pompeu Fabra, Dr Aiguader 80, 08003 Barcelona, Spain.
e-mail: ricard.sole@upf.edu

1. Foley, J. A. *et al. Ecosystems* **6,** 524–532 (2003).
2. Rietkerk, M. *et al. Science* **305,** 1926–1929 (2004).
3. Kéfi, S. *et al. Theor. Popul. Biol.* **71,** 367–379 (2007).
4. Scanlon, T. M., Caylor, K. K., Levin, S. A. & Rodriguez-Iturbe, I. *Nature* **449,** 209–212 (2007).
5. Kéfi, S. *et al. Nature* **449,** 213–217 (2007).
6. Solé, R. V. & Bascompte, J. *Self-Organization in Complex Ecosystems* (Princeton Univ. Press, 2006).
7. Pascual, M. *et al. Phil. Trans. R. Soc. Lond. B* **357,** 657–666 (2002).
8. Marro, J. & Dickman, R. *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge Univ. Press, 2007).
9. Hinrichsen, H. *Physica A* **369,** 1–28 (2006).

ATOMIC PHYSICS

A whiff of antimatter soup

Clifford M. Surko

A molecule consisting of two electrons and two anti-electrons is similar to, but different from, the familiar hydrogen molecule H_2. Its creation heralds a new chapter in the formation of matter–antimatter states.

Particles of antimatter might be rare, fleeting and seemingly unwelcome guests in our matter-dominated world, but they offer many opportunities to study new science and develop new technologies. Antimatter, and how the laws of physics apply to it, is therefore of fundamental interest, notwithstanding the challenges of making, manipulating and storing the stuff. On page 195 of this issue[1], Cassidy and Mills report the breaking of new ground — the creation of the first-ever molecule of a species of matter–antimatter atom known as positronium.

The laws of physics, as we understand them, are symmetrical: for each type of ordinary-matter particle there is a corresponding antiparticle. The proton has the negatively charged antiproton; the electron has the positively charged positron. Such pairs of particles and antiparticles are now regularly created in laboratories around the world. But almost as soon as they are made, they disappear again with a puff and a flash of light, annihilating each other to leave only a trace of other particles or photons. The electron and positron, for example, annihilate into two or three photons with a total energy of 1,022 kiloelectronvolts (keV) — twice the electron mass, 511 keV. These photons are nothing other than highly energetic X-rays or γ-rays, and are routinely used to characterize materials for high-speed electronics, as well as to study metabolic activity in the brain in the technique known as positron emission tomography (PET).

Just as the electron and proton bind to form atomic hydrogen (H), so the electron and positron bind, albeit fleetingly, to form a positronium atom (Ps). The existence of positronium was predicted[2] in 1946 by the theoretical physicist John Wheeler, and the atom was first isolated experimentally[3] by Martin Deutsch in 1951. Wheeler also posited the existence of a dipositronium molecule, Ps_2, and even a triatomic variant, Ps_3. It is Ps_2 that Cassidy and Mills have only now succeeded in producing[1] — and with it the first many-positron, many-electron system to be made in the laboratory.

It is tempting to view Ps_2 as a diatomic molecule similar to the hydrogen molecule, H_2, but there are important differences. Unlike the proton and electron in a hydrogen atom, the positron and electron in positronium have the same small mass of 511 keV. As a result of the quantum uncertainty principle, neither the electrons nor the positrons in Ps_2 can be localized in the same way as the much heavier protons in H_2. Thus, each of dipositronium's four particles has one repulsive partner (of the same type) and two attractive partners (of the opposite type). The four do a merry dance around each other in a fuzzy, lumpless soup with matter and antimatter flavours[4].

Cassidy and Mills performed their experiments[1] by accumulating some 20 million positrons in a specially designed trap. They focused these positrons in a burst lasting less than a nanosecond onto a small spot on the surface of a porous silica sample. The positrons diffuse into voids in the silica, where they capture electrons to form positronium atoms. Before these atoms can annihilate, they form about 100,000 Ps_2 molecules on the interior surfaces of the voids. The presence of a surface is crucial — because the energy of the bound

50 YEARS AGO

"British public schools and the future" — [Another] demand which parents make on the public schools is impossible to justify on educational grounds and has social, political and moral implications. Many parents... send their sons to public schools because membership of these schools will be of service to them in their future careers. Although this charge may be exaggerated — the full effects of the establishment of grammar schools under the Education Act of 1902 have not yet been seen — the public school system undoubtedly confers advantages on its products which are denied to those from State schools. A system which enables less able men to come to the top and prevents the abler from doing so cannot be justified on human, economic or moral grounds. But even this does not provide a case for abolishing schools which have so much to commend them educationally; if the wrong boys are getting into public schools, other means of selecting them must be found.

From *Nature* 14 September 1957.

100 YEARS AGO

Theoretically at least most observers admit that the adoption of the scientific method in the management of the affairs of State is a preliminary necessity if national efficiency is to be secured... There is growing evidence, also, that politicians in most countries are beginning to realise that statesmanship is no exception to this rule, but, like other skilled labour, is most satisfactory when conducted on scientific principles. But whether British statesmen appreciate this truth to the same extent as those of other great nations is a matter of grave doubt. Their education generally has been of such a character as to leave them with a colossal ignorance of science and scientific methods; and it is only by overcoming the bias received at the public school and university that most of them come to understand the modern outlook.

From *Nature* 12 September 1907.

50 & 100 YEARS AGO

EVOLUTIONARY GENETICS

You are what you ate

It is hard to think of anyone who doesn't like starchy foods such as pasta, chips, rice or bread. But certain populations, for example hunter-gatherers living in the rainforests or near the Arctic circle, have historically existed on a diet rich in protein and low in starch. George Perry and colleagues conclude that such differences in the amount of dietary starch have moulded the human genome over time (G. H. Perry *et al. Nature Genet.* doi:10.1038/ng2123; 2007).

Dietary shifts — whether driven by the development of stone tools, by controlling fire or by domesticating plants and animals — have had a major role in human evolution. Perry and colleagues specifically looked at the effect of dietary starch on the number of copies of *AMY1*, the gene that encodes the salivary amylase enzyme, which breaks down starch.

AMY1 is one of the few genes in the human genome that show extensive copy-number variation between individuals. So the authors first looked at whether additional *AMY1* copies are functional. They found that extra *AMY1* copies do indeed endow the individuals carrying them with the capacity to produce more salivary amylase. The question then was whether the starch content of past diets dictates the present levels of amylase and, thus, *AMY1* copy number.

Perry *et al.* studied two groups: one consisted of four populations with a low-starch diet and the other of three populations from agricultural societies and hunter-gatherers in arid environments, who traditionally eat high-starch food.

Strikingly, twice as many members of the high-starch-diet group had at least six copies of *AMY1*. This difference could not be explained by geographical factors because both groups contained people of Asian and African origin. Instead, the authors propose that variations in *AMY1* copy number are more likely to have been influenced by positive natural selection.

So what is the advantage of having more salivary amylase? Significant digestion of starch occurs during chewing. This is crucial, and probably vital, in people likely to suffer from diarrhoeal diseases. Moreover, after being swallowed, salivary amylase is carried to the stomach and intestines, where it aids other digestive enzymes.

Of the three copies of the *AMY1* gene registered in the reference sequence of the human genome, variations in nucleotide

sequences are small. This suggests that the duplication of these genes may have occurred relatively recently, possibly even since the evolution of modern humans about 200,000 years ago. So Perry and colleagues' results, and elucidation of copy-number variations in other human genes, could provide insight into our ecological and evolutionary history.

Sadaf Shadan

Ps_2 molecule is less than that of two Ps atoms moving separately, a third body is required to mop up the excess as recoil energy. The large surface area of Cassidy and Mills' porous silica trap thus provides an ideal place for the two positronium atoms to marry, with the surfaces mopping up the extra energy (Fig. 1).

Evidence for Ps_2 formation comes from studying the temperature dependence of the 511-keV γ-rays that signal the annihilation. Molecules of positronium are formed from a slightly longer-lived version of the Ps atom in which electron and positron spins are aligned. When a molecule is created, a positron can grab an opposite-spin partner, and it thus annihilates more quickly. At lower temperatures, more Ps_2 forms because the atoms are less energetic and stick more readily to the void surfaces. The authors observe an increase in the fast component of the γ-ray signal at lower temperatures, which they interpret convincingly as evidence that molecules are being formed.

The success of this experiment[1], and particularly of the techniques the authors and their colleagues developed to achieve it[5], holds much promise for creating other novel states of matter and antimatter. The current experiment corresponds to a Ps_2 density of about 10^{15} cm^{-3} in the voids. Advances in positron technology should permit the creation of much higher density collections of electrons and positrons[6], allowing us to ask whether Ps_3 or more complicated positronium materials in Wheeler's progression[2] exist and, if so, what their nature is.

In addition, if one could increase the density of positronium atoms to 10^{18} cm^{-3}, the system is expected[7] to undergo a transition at a temperature of 15 kelvin to a 'Bose–Einstein condensate', in which all atoms share the same quantum state. At even higher densities, one might expect the material to become a regular, crystalline solid. If one could make a Bose–Einstein condensate of Ps_2 molecules at densities of 10^{21} cm^{-3}, there is the possibility of creating a laser using the γ-rays from the annihilation process. Because of their high energy, these photons have a very short wavelength, and could be used to probe objects as small as atomic nuclei.

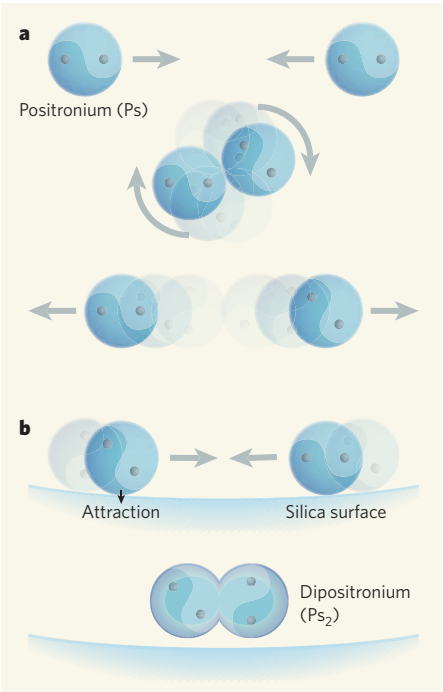


Figure 1 | Positronium gets it together. a, In free space, two atoms of positronium (each consisting of an electron and its antiparticle, the positron) cannot combine to form a molecule — owing to their excess of energy, they simply fly apart again. **b,** A third body, by contrast, such as Cassidy and Mills' silica walls[1], can absorb this excess energy, allowing the two atoms to wed — albeit fleetingly.

Cassidy and Mills' work[1] thus heralds a new front in the study of antimatter and matter. In complementary work, scientists are asking questions about the chemical physics of antimatter — such as the binding of positrons to ordinary atoms[8] — and using the answers to develop new ways to study materials[9]. They are also coming closer to fundamental tests of the symmetry of antimatter and matter by comparing the properties of hydrogen and antihydrogen[10,11]. The pay-off of such tests could be an answer to the perplexing but vital question of why we are surrounded by matter, with so little evidence of antimatter in the Universe. ■

Clifford M. Surko is in the Department of Physics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0354, USA.
e-mail: csurko@ucsd.edu

1. Cassidy, D. B. & Mills, A. P. Jr *Nature* **449,** 195–197 (2007).
2. Wheeler, J. A. *Ann. NY Acad. Sci.* **48,** 219–238 (1946).
3. Deutsch, M. *Phys. Rev.* **82,** 455–456 (1951).
4. Schrader, D. M. *Phys. Rev. Lett.* **92,** 043401 (2004).
5. Cassidy, D. B. *et al. Phys. Rev. Lett.* **95,** 195006 (2005).
6. Surko, C. M. & Greaves, R. G. *Phys. Plasmas* **11,** 2333–2348 (2004).
7. Platzman, P. M. & Mills, A. P. Jr *Phys. Rev. B* **49,** 454–458 (1994).
8. Mitroy, J., Bromley, M. W. J. & Ryzhikh, G. G. *J. Phys. B* **35,** R81–R116 (2002).
9. Ball, P. *Nature* **412,** 764 (2001).
10. Amoretti, M. *et al. Nature* **419,** 456–459 (2002).
11. Gabrielse, G. *et al. Phys. Rev. Lett.* **89,** 213401 (2002).

OBITUARY

Ernst Otto Fischer (1918–2007)

Organometallic chemist, and cosmopolitan Bavarian patriot.

Ernst Otto Fischer belonged to the generation robbed of its youth by the Nazi regime. His bitter experiences on the Russian front in the Second World War had a powerful influence on the liberal and cosmopolitan sentiments of a scientist who dedicated his career to fundamental research. He died aged 88 on 23 July in Munich, where he had lived and worked all his life.

Following his return from war, Fischer studied the chemistry of organometallic complexes under Walter Hieber at Munich's Technische Universität. He and his fellow students had initially to rebuild the badly bombed institute with their own hands. Fischer earned his doctorate there in 1952, with an experimental thesis on a simple synthesis process for the versatile reagent tetracarbonylnickel.

But the spur to become a research chemist came from his father, a physics professor, who drew his attention to a 1951 article in *Nature* on a new type of organo-iron compound known as ferrocene (or 'dicyclopentadienyl iron'). Fischer was sceptical of the bivalency of iron proposed in the paper to explain the compound's structure. Together with Wolfgang Pfab, he used X-ray diffraction to determine its true structure, in which two five-sided carbon rings sandwich a core of a single iron atom. This forms an especially stable compound, with an electronic structure similar to that of a noble gas.

The British chemist Geoffrey Wilkinson had independently come across the same sandwich-compound structure for ferrocene. These were moments of glory that heralded a renaissance of inorganic chemistry. There has been a steady stream of compounds combining metals and organic ligands ever since, with novel structures, reactivities and applications in catalysis. Fischer followed his initial achievement soon afterwards (in 1955) with the discovery of a similar structure, dibenzenechromium, that comprised two six-sided carbon rings sandwiching a chromium atom. This discovery was based purely on theoretical considerations and was made together with his extraordinarily gifted student Walter Hafner.

In the face of strong and not always harmonious competition from Wilkinson (who was initially based at Harvard University, and then at Imperial College in London), Fischer's Munich laboratories produced a host of new organometallic compounds in the years that followed. These included the first compounds with metal–carbon double bonds ('carbenes', discovered together with Alfred Maasböl in 1964) and metal–carbon triple bonds ('carbynes', with Gerhard Kreis in 1973). Whereas Wilkinson became interested at an early stage in the catalytic effects of these organometallic materials (such as the use of rhodium complexes to catalyse hydrogenation and oxosynthesis reactions), Fischer focused exclusively on investigating the rich structures and reactivities of the new world of organometallic chemistry. The two shared the Nobel Prize in Chemistry in 1973.

The long-term influence of Fischer's extensive experimental work is impressive, and has opened up new horizons, also extending to practical applications. The best example is the technique of 'olefin metathesis', in which bonds between organic building-blocks are redistributed to produce new products useful in medicine and industry. The catalysis cycle of this process involves the formation of a metal–carbene intermediate. Without Fischer's original research, this advance — which culminated in the Nobel prize awarded to Yves Chauvin, Robert Grubbs and Richard Schrock in 2005 — would not have been possible.

Throughout his life, Ernst Otto Fischer remained deeply attached to his native Bavaria and his home city of Munich. Following his first, brilliant individual research results on ferrocene and similar organometallic complexes, he moved in 1957 to the city's Ludwig-Maximilians-Universität. In 1964, he returned to take the chair of his own mentor, Walter Hieber, at the Technische Universität, a position he held for 20 fruitful years.

Fischer's special talent was to radiate delight in all things new, a trait that served to motivate and inspire his charges. A bachelor throughout his life, he regarded the members of his research group as his family. Once he had established their competence — and he had an unerring instinct for young talent — he granted them unlimited scope to unleash their creativity. In this way, all the researchers who clustered around him in Munich, more than 200 in total, felt a personal responsibility for the common cause.

Shaped in this way, Fischer's students became role models in their turn, shouldering responsibility for the community, passing on knowledge and creating new levels of accountability among an ever-expanding group of researchers. More than a dozen of his postdoctoral students were awarded chair professorships at German universities, and many of his alumni later rose to the higher echelons of the chemical industry.

The contributions of Fischer's protégés to basic and industrial research has been just as significant. A particularly fitting example is the Wacker process to produce acetaldehyde. This uses an ingeniously simple catalytic conversion technique based on ethylene, oxygen and water, and was devised by Walter Hafner together with another of Fischer's former students, Reinhard Jira.

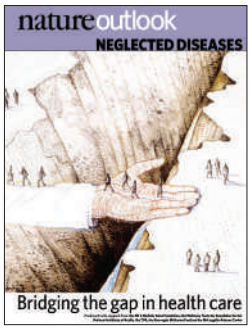
Beyond the realm of chemistry, Ernst Otto Fischer was a highly erudite man who captivated his international colleagues with his profound knowledge of literature and music. He could be both impulsive and contemplative, emollient and combative, and coupled a love of his country with a cosmopolitan outlook — seemingly contradictory traits that found their human symbiosis in him. For colleagues such as myself, he was the epitome of the contentious, relentless fighter for the freedom of science, who fiercely rejected any manipulative outside influences. He was the grand old man of fundamental research, who would settle for nothing less than creativity and scientific curiosity.

With his death, his Bavarian homeland has lost one of its most illustrious sons. He found his final resting place in the Old Cemetery in Munich-Solln on 26 July 2007.

Wolfgang A. Herrmann
Wolfgang A. Herrmann is in the Department of Inorganic Chemistry and is president of the Technische Universität München, Arcisstraße 21, 80333 München, Germany.
e-mail: praesident@tum.de

nature outlook

NEGLECTED DISEASES



A. E. COBER/IMAGES.COM/CORBIS

"We have never had such a sophisticated arsenal of technologies for treating disease, yet the gaps in health outcomes keep getting wider. This is unacceptable." This plea to close the gap between rich and poor nations was made last month by Margaret Chan, director-general of the World Health Organization (WHO), in her first major address on primary health care. Few would disagree. The tragedy is that it joins a litany of similar unheeded appeals by WHO directors-general, stretching back almost 30 years.

Today, more than one billion people — one-sixth of the world's population — are affected by tropical diseases, yet the drug 'pipeline' for these neglected diseases is almost dry. Clearly, there is an urgent need to develop and deliver effective new therapies.

There is no shortage of high-quality research into tropical diseases. But the successful translation of basic scientific discoveries into new treatments is a long, complex and expensive process. This is particularly problematic when dealing with diseases that disproportionally affect poor and marginalized populations. Low returns on investment have discouraged drug companies from allocating resources to tackle these diseases. And an academic culture that rewards publications and wealth creation, rather than contributions to practical social good, has exacerbated the problem and widened the 'translational gap'.

Despite recent increases in funding for developing and delivering new drugs, the situation has not noticeably improved for those actively engaged in the fight against neglected diseases. This is because the problem cannot be solved solely by increasing financial resources. Improvements in infrastructure are sorely needed at almost every level, from regulatory-authority involvement to government leadership and research capacity. As is clear from the following pages, bringing scientific innovation out of the laboratory and into the villages is not a simple task, and success will require the cooperation and commitment of numerous institutions, both public and private. Preventing the millions of unnecessary deaths that occur each year from neglected diseases is the goal, and anything less is unacceptable.

We are pleased to acknowledge financial support from the Bill & Melinda Gates Foundation, the Wellcome Trust, the Foundation for the National Institutes of Health, the Special Programme for Research and Training in Tropical Diseases (TDR), the Burroughs Wellcome Fund and the McLaughlin-Rotman Centre for Global Health in producing this Outlook. As always, *Nature* carries sole responsibility for all editorial content.

David O'Connell
Chief Editor, *Nature Reviews Microbiology*


p170

features

p160

Lost in translation

The culture of academia needs to change if scientists are to bridge the gap between research and the development of drugs and vaccines for neglected diseases in the developing world, says **Declan Butler**.

M. HENLEY/PANOS

Many promising research advances never make it to the people who need them most.

Read any scientific paper or grant proposal on basic research into neglected diseases and you inevitably find a claim that the work could lead to new therapies for diseases that affect millions of people in the developing world. Few, in fact, ever do, and scientists, universities and research funders are waking up to the fact that part of the blame lies with them and their perpetuation of a reward culture that focuses excessively on papers and patents, and not on whether the research actually benefits society.

In disease research, academia has traditionally restricted its role to basic research. Subsequent development is then left to the pharmaceutical industry. But when it comes to neglected diseases — those that disproportionally affect poor and marginalized populations — the drugs and vaccines have low returns, so commercial firms cannot fork out for the expensive development. As a result, there is a 'translational gap' in which promising research leads sit on the shelf, and potential drugs and vaccines go undeveloped.

The academic reward culture means that researchers have little incentive to do the 'grunt work' needed to close this gap because it does not generate papers in top journals, says Mel Spigelman, director of research and development at the Global Alliance for TB Drug Development in New York. The competitive nature of academic research and of the publication reward system can also lead to unnecessary duplication of work, and hinders the sharing of data, he adds.

> "Having a chemical structure that kills your parasite is only one of many aspects of what makes a drug a drug."
> — Els Torreele

The time has never been riper for greater engagement of academics in the development of new drugs. Over the past decade, the translational gap has started to close for neglected diseases, thanks to the emergence of a series of public–private partnerships between charities, large pharmaceutical companies and small biotech firms. These partnerships are run like businesses, but not for profit, and include the Malaria Vaccine Initiative in Bethesda, Maryland; the Medicines for Malaria Venture in Geneva, Switzerland; the Global Alliance for TB Drug Development; and the Drugs for Neglected Diseases Initiative (DNDi) in Geneva, Switzerland, which is seeking cures for the 'most neglected' diseases, such as sleeping sickness (African trypanosomiasis), Chagas' disease and leishmaniasis.

Although only 13 of the 1,223 drugs developed since 1975 have been for neglected diseases, dozens are now in the pipeline of public–private partnerships. These partnerships also play a key part in helping academics to tailor their research to the needs of drug development by creating focused projects that bring them into contact with industry scientists (see pages 164 and 166).

That cross-fertilization is important because translational research requires skills and a culture that universities typically lack, says Victoria Hale, chief executive of the non-profit drug company the Institute for OneWorld Health in San Francisco, California, which is developing drugs for visceral leishmaniasis, malaria and Chagas' disease. Academic institutions are often naive about what it takes to develop a drug, she says, and much basic research is therefore unusable. That's because few universities are willing to support the medicinal chemistry research needed to verify from the outset that a compound will not be a dead end in terms of drug development.

Academics will currently publish, say, a chemical scaffold, which they bill as a potential new target for parasites. "But had a medicinal chemist looked at it, he might immediately see that it will never work as a drug, because it has an inappropriate solubility or toxicological profile," says Els Torreele, a product manager at the DNDi. "Having a chemical structure that

kills your parasite is only one of many aspects of what makes a drug a drug."

Ted Bianco, director of technology transfer at the Wellcome Trust in London, agrees. "It's fine if a researcher is just using a compound as a ligand to probe a biological process," he says, "but don't kid yourself it's a drug unless you ask whether it has druggable properties." What's needed, says Hale, is a 'target product profile', which sets out the appropriate drug chemistry properties. "Getting a drug through regulatory processes is not just about how good your science is and how great your trials are; it is much more complex," says Hale. "And academics don't have the experience — they need to hire people from the drug industry."

> **"Non-patenting or free licensing might be the best way to make sure that research is used for public benefit."**
> — Carol Mimura

Seeds of success

To tackle this problem, the Wellcome Trust has started a £91-million (US$182-million) five-year Seeding Drug Discovery initiative for neglected diseases, which aims to provide universities with the funds to do the sort of screening and medicinal chemistry usually found only in industry. The initiative includes an £8.1-million award to the University of Dundee, Scotland, which is working with the DNDi on drugs for the most neglected diseases.

But universities could also play a part by engaging actively in development research for neglected diseases, says Hale. Over the past two decades, academic institutions have been putting greater emphasis on wealth creation. Universities will often patent everything they can in the hope of generating additional revenue streams. The thicket of resultant patents has meant that groups seeking to develop drugs and vaccines for neglected diseases simply cannot afford to do the necessary research.

Although universities with strong tropical-disease research centres are often well aware of the need to make patents freely available, many techniques needed in drug development, such as drug stabilization, are patented by universities that have no interest in tropical diseases. Persuading a university to let organizations such as the DNDi use a technology freely often takes more than a year, and almost always requires intervention at the highest levels of the university, says Torreele.

The University of California, Berkeley, has been a pioneer for change here. In 2004, its technology-transfer office was approached by Eva Harris, a researcher at the university's school of public health. Harris wanted to create a project to field-test a cheap, hand-held, electromechanical device to diagnose dengue fever. The snag was that the non-profit project couldn't pan out if it had to pay royalties on the underlying patents, held by the university.

The idea of handing out royalty-free licences was radical at the time, recalls Carol Mimura, assistant vice-chancellor at the Office of Intellectual Property & Industry Research Alliances at Berkeley. The university not only agreed to Harris's request, but built on it, and last year decided to free up patents for humanitarian uses in its socially responsible licensing programme, which is headed by Mimura. The programme has since spawned a dozen or so projects, including a cheaper artemisinin drug for malaria, and a possible vaccine for tuberculosis (see page 174).

Harris provided a moral compass that led to a cultural shift at Berkeley, says Mimura. "Organizations perform to what they are measured by. Traditional metrics such as the numbers of patents, licences signed and amount of licensing revenue, seem to be the sole goals that a technology-transfer office aspires to," she says, "when in fact non-patenting or free licensing might be the best way to make sure that research is used for public benefit."

Although it is more difficult to measure, Berkeley now treats the social impact of its intellectual property as an equal or more important bottom line than economic gain, says Mimura. She thinks that the university will gain in the long term by attracting donations and philanthropic partnerships as a result of the reputation and goodwill generated. "We need to manage our intellectual property with an eye towards the ultimate ripple effect of its management, not just the short-term outcomes," she says.

Pressure tactics

In 2001, Yale University and Bristol-Myers Squibb agreed to allow South Africa to make generic versions of an AIDS drug, d4T (stavudine), which Yale had licensed to the company. Pressure has also come from a student group based in Philadelphia called Universities Allied for Essential Medicines, which last November released the Philadelphia Consensus Statement, which called on universities to adopt policies in favour of neglected diseases.

And the momentum is growing. The US National Institutes of Health has adopted similar progressive policies to those of Berkeley, as have several universities in North America. And in March, a dozen universities in the United States issued a joint statement pledging to be "mindful of their primary mission to use patents to promote technology development for the benefit of society", including making provisions for therapeutics, diagnostics and agricultural technologies in developing countries (see http://tinyurl.com/2qmtr4).

But if the pendulum is swinging back from an emphasis on wealth creation to one on social good in the United States, the same is not true everywhere, says Torreele. Ironically, it is now often easier to get royalty-free patents from a US university than it is from some in India, Brazil or Thailand, she says, as emerging economies follow the West's lead in emphasizing wealth creation from academic research. ■

Declan Butler is *Nature*'s European correspondent.

I. GETACHEW/UNICEF ETHIOPIA

Vaccination programmes in countries such as Ethiopia depend on having affordable products.

A tough transition

What is holding back biotechnology in the developing world? **Peter A. Singer** and his colleagues listen to those on the ground.

P. SINGER

The path from basic scientific discovery to effective therapy is rarely rapid or simple, especially in the developing world. Making this transition easier is a sizeable and pressing problem. What is the best way to tackle such a complex issue?

One important step is to identify the factors that help and hinder the uptake of health-related biotechnology in developing countries. Although knowledge about these factors in developing regions is lacking[1], the spread of technologies and ideas across cultures has been studied extensively for several decades. The complex issues involved in the development of new technologies cover areas as diverse as science-capacity building[2], culture[3], economic analysis[4], foreign investment and imports[5], public–private product-development partnerships[6], intellectual property[7] and political policy[8]. These issues, however, have mostly been explored in the context of the developed world and in isolation from one another, which means that the bigger picture remains unclear.

The diversity of factors involved in the developing world is illustrated by several recent events. Treatments for HIV infection are urgently needed in Africa, yet clinical trials of the anti-HIV drug tenofovir were halted in Cambodia, Cameroon and Nigeria because of claims that efforts to inform and involve local communities were inadequate[9]. And, in the face of famine, the Zambian government rejected food donated by the United States because it was genetically modified[10]. On a longer timescale, there was a large delay between the development of a vaccine against hepatitis B in the developed world and its widespread availability in the developing world.

At present, there is great interest in exploring such issues because of the unprecedented increase in financial resources for improving the health of individuals in the world's poorest countries. For example, since 2005, biomedical research projects focusing on the developing world have received US$450 million from the Grand Challenges in Global Health initiative of the Bill & Melinda Gates Foundation. With the increase in resources comes added responsibility to ensure that optimal improvements are made.

We set out to identify the factors, or forces, that affect the uptake of health-related biotechnology in the developing world. To do so, we interviewed 70 key experts from various sectors — academia, industry, civil society (voluntary and civic organizations) and government — in developing countries (see Supplementary Information for full details and breakdown).

These interviews allowed us to identify eight key areas that affect the development and adoption of health-related biotechnology in resource-poor regions. These areas can be categorized into four main forces: scientific, social (including ethical and cultural), financial and political. As a result, we were able to generate a model that can be used to assess the likelihood of success of health-related biotechnologies (see graphic, overleaf).

The scientific question

In terms of science, the participants in our study highlighted scientific capacity, infrastructure and collaboration as important.

Inadequate scientific capacity and infrastructure — for research, manufacturing and delivery — slow the development and adoption of health-related biotechnologies. Our interviewees emphasized that building the capacity for scientific research in the developing world — for example, improving training, facilities and equipment — enables researchers to investigate neglected diseases. But even for nations that have a strong research record, such as South Africa, the capacity for manufacturing remains a bottleneck. Local development and manufacturing, where possible, help local economies, making new technologies more sustainable and acceptable.

In the developing world, especially in rural areas, poor infrastructure also hampers the distribution of medicines. For example, understaffed health clinics need to cope with unreliable or prohibitively expensive transport. They often have no electricity or are plagued by power cuts. And they can lack potable water for administering pills.

Scientific collaboration can improve local scientific capacity, potentially enabling the developing world to contribute to the development of new technologies. But collaborations set up between scientists in the developing world and those in the developed world must be defined at the outset to avoid potential conflicts or exploitation.

One-third of all interviewees recommended classifying collaborations between scientists in the Northern Hemisphere and Southern Hemisphere as 'south–north' to remove any connotation of northern dominance. In addition, south–south collaborations hold promise for developing regionally relevant health-related biotechnologies and for creating sustainable benefits in the developing world. But care must be taken to avoid poaching scientific experts from other southern countries.

Social issues

When it comes to social issues, engaging the local community or general public and understanding the acceptability of products to the local culture seem to be key.

Engaging communities involves both authorization and consent. It depends on a legitimate and democratic leadership, and the acceptance of, and involvement in, new research strategies by the community. Our interviewees highlighted the need for early community engagement in any negotiations and deliberations, to help relationships between investigators and communities to be successful. Furthermore, they indicated that

Act local: drugs made in Ghana are cheaper and more acceptable to the local community.

informed consent and community ownership, as well as access of participants to health care while enrolled in the trial and to successful drugs after the trial, are crucial to prevent abuse of the community.

Engaging the public helps to overcome the social barriers against new technologies such as nutritionally enhanced foods, intravaginal vaccines and genetic strategies to control mosquitoes. Information from radio, television and print media can help members of the public to evaluate their options. Similarly, educational campaigns led by religious leaders, political champions, tribal chiefs or teachers can enable people to make informed choices.

On the basis of previous vaccination and reproductive-health campaigns, the adoption of certain new technologies, including several being developed by the Grand Challenges in Global Health initiative, will require a well-designed strategy for engaging the public, gauging their views and concerns, and addressing these when implementing the intervention.

The cultural acceptability of new technologies must not be neglected, otherwise they will not reach the people who need them. To raise cultural awareness, researchers should allocate resources to the understanding of cultural diversity. This diversity is rooted in complex, interrelated cultural issues that stem from gender, religion, historical context, sexual practices and the use of contraception, and the absence of a culture of science in some countries. But as Musimbi Kanyoro, then general secretary of the World YWCA, said, culturally derived moral positions themselves have limits: "How can we, at this moment, let morality override mortality when we see the numbers of dead?"

Financial considerations

When it comes to finance, what is really meant is affordability and commercialization. New medicines developed as a result of advances in biotechnology must be affordable to people in developing countries (see page 176), and new drugs should be commercially viable in low-resource settings.

The commercialization of products for developing countries might entail financial risk. And entrepreneurs must take into account the limited funds available for production and the small profit margins in these regions. One way to boost commercialization is to manufacture drugs locally, making these drugs more affordable.

Increased affordability, in turn, results in access to larger markets and more opportunities

LOCAL VOICES

People working in developing countries offer their take on what affects the success of biotechnology in these regions.

"You can't even begin to think about vaccines until you've succeeded in delivering electricity that will run refrigerators or have solar-powered refrigerators or something in the area to keep the vaccine cold."

"What we want Western participation for is to bring innovation into thinking here — not to think for us."

"You may end up with this fantastic, efficacious magic bullet, but if you don't understand some of the local culture and customs etcetera, then the uptake and acceptability are low."

"There's a big division between what the needs are in these countries and where drug development or other biotechnologies are actually located and shaped."

"If a particular technology is very expensive, then the chance that it will be adopted and solve a particular problem is likely to be low."

"The principle of justice demands that if a technology is found to be effective it should be made affordable to the population on whom it was experimented."

"In typical African culture, men are always perceived as superior, and as a result, women do not have a voice."

"Researchers will continue to work hard to develop an efficacious product, but no product will reach the people who need it if the policy-makers won't support it."

"India provides an enormous cost advantage in process and clinical drug development. Multinational pharmaceutical companies have been forced to substantially reduce their prices."

"Western scientists come with ready-made plans, and these don't necessarily answer local questions."

"It comes back to political will and guts. Someone has to step up there, be the champion and say 'Come hell or high water, this is what we're going to do'."

"There's a danger that much of this research that is being done on technology will result in nothing because we have not understood what the community needs are."

for industry to balance profit and social benefit. An example of successful commercialization is India's first domestically produced and marketed recombinant vaccine against hepatitis B. The innovative manufacturing processes used by Shantha Biotechnics, in Hyderabad, made this vaccine affordable to the Indian population[11]. Another example is the vaccine against human rabies produced by Indian Immunologicals, also based in Hyderabad. The company developed a distribution network of refrigerated vehicles and franchise clinics to ensure that the vaccine reaches rural villages[11].

Another crucial challenge for industry in the developing world is the weakness or absence of regulatory systems, which stifles innovation and patent registration and undermines the public's trust. Furthermore, our interviewees acknowledged that respect for intellectual property will be crucial if countries are to join the 'development ladder' (see page 174). But they emphasized that addressing global health crises will require ingenuity in devising commercialization strategies and guarding intellectual property.

For many interviewees, the issue of affordability was underpinned by considerations of fairness. Interviewees saw clear roles and responsibilities for providing affordable technologies to the poor on the part of the private sector, the government and international organizations. These involve differential pricing (supplying products to different markets at different prices) and single-buyer markets (for example, governments or the World Health Organization), subsidies for vaccines and drugs, building up infrastructure and providing health care. The moral imperative to make products affordable is especially strong when considering those who participate in successful clinical trials but cannot afford to continue treatment when the trial ends. Overall, the main challenge in terms of finances is to balance affordability with incentives to innovate.

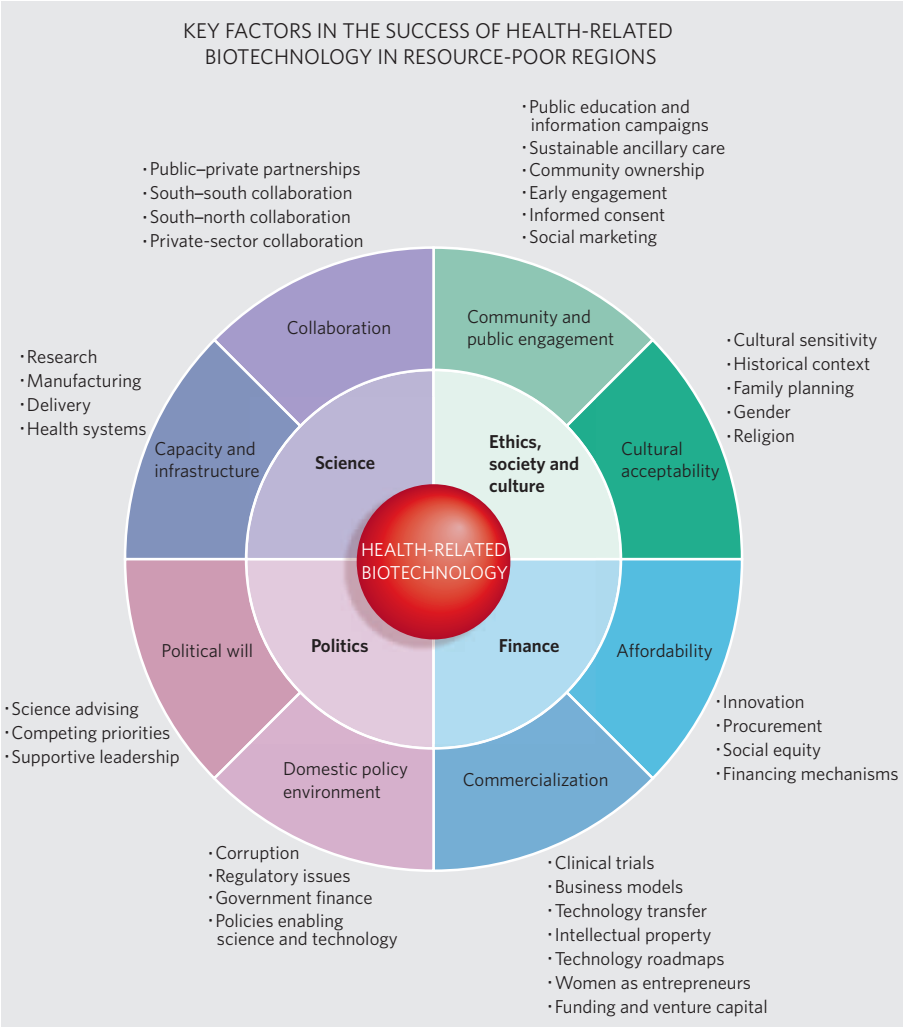
Politics and policy

The final area that affects the development and uptake of health-related biotechnology is politics: political will and the domestic policy environment. Our interviewees stressed that governments should include biotechnology in policies aimed at broadening national development and that these policies should span the sectors of medicine, health, development and trade.

Governments have the potential to fund the development and manufacture of drugs and to supply the single-buyer markets that make drugs affordable. Our interviewees called for governments in the developing world to follow the example of countries such as Brazil, China, Cuba, India, Rwanda, South Africa and Thailand, which have reaped the benefits of policies giving priority to science and technology. For example, the world's first vaccine against infection with *Neisseria meningitidis* serogroup B, which causes meningitis[12], was produced in Cuba. Antigens manufactured in Cuba are also distributed to other developing countries: for

A. NATH/AP

Engaging the local community is crucial to the success of new ideas.



example, for the hepatitis B vaccine produced at South Africa's Biovac Institute in Wadeville.

Corruption, competing political priorities, inadequate public infrastructure for health care and anti-biotechnology lobbies can all block funding and access to new health products. Affordable hepatitis B vaccines, for example, fail to reach South Africans who need them, partly because public-health programmes are not politically expedient. Strong political champions, however, can forge progress. This has been shown by the New Partnership for Africa's Development, which has increased the visibility of science and innovation to African heads of state and has acknowledged the link between economic transformation and innovation in science and technology.

A model for the future

Several of those interviewed for our study stressed the importance of a broad approach to health-related biotechnology issues. For example, Adi Paterson, Group Executive of the South African government's Department of Science and Technology, urged all investigators to consider ethical factors in their analysis: "When a model is science-centric, it loses its ability to actually reflect early on the ethical question."

As a result of our interviews, we have assembled a model that can be used to assess the potential success of particular health-related biotechnologies with respect to bottlenecks, regional variations and changes over time. (see graphic, above). In addition to the more obvious issues relating to science and finance, our model highlights that other issues — ethical, social, cultural and political — need to be addressed when seeking to improve health in the developing world.

It provides a framework for taking health-related biotechnology from the lab to the village. It offers a guide to structuring product and market readiness and to identifying factors that could impede the development of new products, such as vaccines, nutritionally enhanced foods, chemical and genetic strategies to control vectors, and diagnostics.

Using the model to assign priorities should streamline the development and adoption of emerging technologies. The model could, for example, be applied to the 2007 report *Freedom to Innovate: Biotechnology in Africa's Development* by the African Union and the New Partnership for Africa's Development, specifically to the 20-year comprehensive approach to fostering health-related biotechnology in the region. In this context, the model could provide a blueprint for commercializing African health products.

We asked our study participants only about health-related biotechnology, so we cannot claim that the model applies to other types of technology, although it might resonate with some. We also acknowledge that the limited number of Latin American participants could diminish the scope of the model's application in that region, so the model might not apply in all developing-world contexts. However, key themes recurred in different regions, supporting the general relevance of the model. Some components of the model might seem simple and self-evident to experts, but this only affirms the validity of our empirically generated model.

Backed by the voices of experts in the developing world, the model enables the complex issue of biotechnology development to be systematically tackled, through examining the forces that affect health-related biotechnology in developing countries. The millions of deaths from preventable diseases every year demand clarity in interpreting and addressing these factors. ■

Peter A. Singer, Kathryn Berndtson, C. Shawn Tracy, Emma R. M. Cohen, Hassan Masum and Abdallah S. Daar are in the Program on Life Sciences, Ethics and Policy, at the McLaughlin–Rotman Centre for Global Health, Toronto, Canada. James V. Lavery is at the Centre for Research on Inner City Health and the Centre for Global Health Research, at St Michael's Hospital, Toronto, Canada.

1. Daar, A. S. *et al. Nature Genet.* **32,** 229–232 (2002).
2. Morel, C. *et al. Science* **309,** 401–404 (2005).
3. Maitland, C. *AI Soc.* **13,** 341–356 (1999).
4. Geroski, P. A. *Res. Pol.* **29,** 603–625 (2000).
5. Keller, W. *J. Econ. Lit.* **XLII,** 752–782 (2004).
6. Widdus, R. & White, K. *Combating Diseases Associated with Poverty: Financing Strategies for Product Development and the Potential Role of Public–Private Partnerships* (Initiative on Public–Private Partnerships for Health, Global Forum for Health Research, Geneva, 2004).
7. Commission on Intellectual Property Rights, Innovation and Public Health. *Public Health Innovation and Intellectual Property Rights* (World Health Organization, Geneva, 2006).
8. Stoneman, P. & Diederen, P. *Econ. J.* **104,** 918–930 (1994).
9. Singh, J. A. & Mills, E. J. *PLoS Med.* **2,** e234 (2005).
10. Mitchell, P. *Nature Biotechnol.* **21,** 6 (2003).
11. Frew, S. E. *et al. Nature Biotechnol.* **25,** 403–417 (2007).
12. Thorsteinsdóttir, H. *et al. Nature Biotechnol.* **22,** DC48–DC52 (2004).

Supplementary information is linked to the online version of this article at www.nature.com/nature

Acknowledgements We thank the study participants; the Ethical, Social and Cultural Program Team, especially A. Bhan and P. Tindana; the Grand Challenges in Global Health initiative staff for comments; and M. Keating for editorial input. This study was funded by a grant from the Bill & Melinda Gates Foundation, through the Grand Challenges in Global Health initiative.

The path to new medicines

Governments must help accelerate the development of drugs needed to treat infectious diseases in the developing world, say **Bénédicte Callan** and **Iain Gillespie**.

RENE VERLEG FOTOGRAFIE

Over the next decade, it should be possible to produce a new generation of safe, effective and inexpensive medicines for many of the infectious diseases that afflict the poor. To achieve this, it will first be necessary to address the lack of viable commercial markets, to scale up the global capacity for research and development (R&D), and to build a more efficient and more open mechanism for the discovery of new drugs. Governments can provide the leadership necessary to align the increasingly political issue of global health with philanthropic funding, technological capability and the new opportunities stemming from scientific progress. These are all increasing steadily and now is the time for governments to act.

In June the Organisation for Economic Co-operation and Development (OECD) held a high-level forum on neglected and emerging infectious diseases in Noordwijk, the Netherlands. It brought together senior representatives from government, industry and academia and from philanthropic, international and non-governmental organizations. They discussed how to build strong international support for accelerating the development and delivery of new medicines, vaccines and diagnostic tests for diseases that disproportionally affect developing countries. The consensus that emerged is summarized in the action points of the Noordwijk Medicines Agenda (see box).

> "The top priority in overhauling the innovation system has to be the delivery of medicines for infectious diseases to the poor."

Although participants made it clear that many health issues in developing countries will not be solved by new technologies alone, these will still be important for reducing poverty and its consequences. The forum called on governments to show political leadership by joining with industry, product-development partnerships (PDPs), investors, shareholders, and intergovernmental and non-governmental organizations to intensify the cooperation and collaborations that will improve access to new health technologies for infectious diseases.

'Push' and 'pull' tools

Several experiments that include 'push' and 'pull' mechanisms have been introduced since 2000 to spur innovation in the fight against infectious diseases. Push mechanisms increase investment in research at the start of the innovation pathway: for example, by subsidizing the costs incurred when developing products for unprofitable or unpredictable markets. The most promising new push mechanisms involve public–private PDPs, which optimize leads, select candidates and bring products through clinical trials. The dozen or so existing PDPs are mainly funded by philanthropic organizations. Other push tools include basic research funding, targeted R&D funds (such as the proposed Industry R&D Facilitation Fund: www.wellcome.ac.uk/assets/wtx026592.pdf) and tax credits.

Pull mechanisms — such as advance market commitments (AMCs), patent extensions, prizes and patent buyouts — are designed to provide incentives for the development and manufacture of usable technologies towards the end of the innovation pathway. They motivate investment by guaranteeing a reward for the product after the completion of its development phase. Pull mechanisms are politically attractive because they address a specific need (for example, lack of a market), are outcome-oriented, and are bounded by time and expense. In theory, pull mechanisms should stimulate a wide variety of discovery efforts in a competitive process but are probably most appropriate when the technological route is marked out. In early 2007, a pilot AMC for the development of a vaccine against pneumococcal disease was launched with a US$1.5-billion commitment by five nations and the Bill & Melinda Gates Foundation, and a malaria-vaccine AMC is being planned. These funds will subsidize the purchase of a vaccine when it has been developed and is in demand in developing countries.

Designed correctly, a combination of push and pull mechanisms — including subsidies and markets guarantees — could facilitate the development of new vaccines and drugs for neglected diseases. We still do not know what the optimal mix of these policies is likely to be. So it is crucial to establish appropriate metrics for evaluating performance, to understand how to tailor different incentives for a broad range of diseases and treatments.

Open innovation networks

To increase the number of industry and public laboratories involved globally in research into neglected infectious diseases and to maximize

June's high-level forum in Noordwijk set out goals for better access to drugs for neglected diseases.

the effectiveness of their contributions, a more fundamental transformation of the innovation pathway is necessary. Fortunately, this transformation process has already begun.

Failures in the innovation system can impede the development of appropriate health technologies for the developing world. These can occur at the level of generating and optimizing leads, during the rational identification and selection of candidate drugs from existing compounds, and in the clinical trials used to test new drugs or regimens. In particular, upstream research and early 'proof of concept' work, which provide new leads and create a pipeline of possible new health products, are weak.

The PDPs overcome these barriers by outsourcing knowledge, compounds and tools. The Special Programme for Research and Training in Tropical Diseases (TDR) is one organization that is developing a virtual drug-discovery capacity by using a series of portfolio, screening and medicinal-chemistry networks.

A more open innovation and collaborative research environment would also increase the efficiency and lower the costs of developing new, safe and effective medicines, vaccines and diagnostics through virtual networks (see page 166). What is needed is a better balance between stimulating innovation and providing broader access to knowledge. There are several tools, such as clearing houses and patent pools, and organizational forms, such as networks and consortia, that would promote easier and more open access to elements such as knowledge, data and process innovation. The challenge is to apply these to the neglected infectious diseases. Specific proposals for doing so include creating a shared global portfolio of prioritized drug-discovery projects and a portal for shared drug-discovery tools; matching potential collaborators on a particular project; supplying privileged access to chemogenomics data; and developing common platforms of intellectual property and management-support services.

Intellectual-property rights

The term 'open' applied to innovation does not necessarily mean a freely available source and the absence of intellectual-property protection. The Noordwijk Medicines Agenda recognizes that the protection and use of intellectual-property rights are important for encouraging investment in R&D, but these might not be sufficient to stimulate innovation as far as the neglected and emerging infectious diseases are concerned. But to attract and expand industry participation in such open networks, intellectual-property rights will need to be respected (see page 174), but the norms could be modified within the network. The intellectual property generated by the virtual teams within an open network is likely to be protected as it is in any other public–private collaboration.

Ideally, standard collaboration agreements would facilitate the rapid formation of collaborative arrangements. One possibility is that intellectual-property rights for any successful drug candidates produced would be licensed or donated to the sponsoring PDPs at the start of the clinical-development programme, although rights could be retained for use in other indications. In short, the network model proposed here is 'open' in the sense that it facilitates broader access to, and use of, data, knowledge and inventions within a worldwide network of researchers, but it would function within the present intellectual-property regime.

Broader innovation benefits

The drug market for infectious diseases in the developing world is both a challenge and an opportunity for the pharmaceutical industry, PDPs and other stakeholders. Unattractive, low-margin markets can be fertile ground for sparking innovation. With lower revenue growth predicted across the pharmaceutical industry in the coming decade, cost containment is becoming an industry-wide problem.

There might be relatively little economic profit to be gained from the development and licensing of drug candidates for the neglected diseases themselves. But there is an economic opportunity in applying the lessons learned from low-cost drug discovery for developing-world diseases to the wider range of niche and segmented non-communicable-disease markets in the developed world. The operation of networks as an emerging model of drug discovery could be an important innovation in its own right.

The top priority in overhauling the innovation system has to be the delivery of medicines for infectious diseases to the poor. Governments can and should use this opportunity to drive a health-innovation strategy that is more efficient and reactive to global public-health needs — one that will leave our innovation systems in better health to deal more effectively with the challenges. ■

Bénédicte Callan is principal administrator of, and Iain Gillespie is head of, the Biotechnology Division at the OECD, Paris, France.

THE NOORDWIJK MEDICINES AGENDA

Recognizing that it is important to scale up and expand new for-profit and non-profit models of innovation for tackling neglected infectious diseases in the developing world, the Noordwijk Medicines Agenda calls for several changes to the present health-innovation system (for full details, see www.oecd.org/sti/biotechnology/nma).

- Prioritize research and development (R&D) needs and align research to a common purpose
- Assess the viability of a global virtual network for drug development that draws on and scales up existing research networks and is more open
- Create incentives for R&D through alternative policy mechanisms to reward innovation
- Facilitate the development and operation of a sustainable architecture for sharing and exchanging knowledge, data and research tools
- Identify the infrastructure necessary for a global virtual collaborative network
- Explore collaborative mechanisms for intellectual-property management
- Promote the transfer of technology, knowledge and technical skills to strengthen innovation systems in developing countries
- Forecast the demand for medical technologies for neglected and emerging infectious diseases
- Support and provide incentives to new for-profit and non-profit models of partnerships between developing and developed nations to accelerate R&D for neglected diseases **B.C. & I.G.**

Mission possible

One billion people worldwide suffer from tropical diseases. **Andrew L. Hopkins**, **Michael J. Witty** and **Solomon Nwaka** explain how drug-discovery networks might be scaled up to address the lack of treatments cost-effectively.

New drugs are urgently needed for neglected tropical diseases and tuberculosis. Although one person in six suffers from such diseases, the few drugs available are not widely used, owing to problems with safety, administration, cost and — increasingly — resistance of the infectious agents[1]. Moreover, there is a shortage of potential drugs for these diseases. So how do we meet this challenge and accelerate the flow of drugs through the pipeline?

The dearth of new drugs entering development for tropical diseases results mainly from the gap between basic scientific research, which is usually publicly funded, and clinical development, which is usually funded by pharmaceutical companies or, more recently for diseases in the developing world, through public–private partnerships. This situation reflects, in part, a lack of commercial incentives to develop new drugs for use in the developing world.

There are several current strategies for tackling this problem. Pull mechanisms, for example, offer a guaranteed market for a product — or other rewards for the company — when the development phase is complete. Push mechanisms involve subsidies to support developing pharmaceutical products for unprofitable or unpredictable markets. Public–private partnerships and non-profit organizations have emerged as cost-efficient vehicles for clinical-product development, and they tend to focus on one or a few neglected diseases. Examples are the Medicines for Malaria Venture, the Institute for OneWorld Health and the Special Programme for Research and Training in Tropical Diseases (TDR).

> "About half of the drugs being developed to treat neglected diseases fail the criteria for being fully effective in the field."

But as most experimental drugs fail in the development phase, the challenge is to produce a sustainable pipeline of new drug candidates to enter development. In the past few years, research facilities to undertake drug discovery for tropical diseases have been established in both industry and academia. A simple calculation, however, shows why the discovery of new drugs for neglected infectious diseases is unrealistic for a company or institute acting alone.

There are at least eight diseases for which new drugs are urgently required, according to the TDR. If a realistic goal is set today to deliver at least one effective new medicine for each of these eight diseases by 2020, then at least six candidate compounds need to enter clinical trials for each disease (assuming a historical success rate of 16% for anti-infective drugs)[2]. At present, in the pharmaceutical industry, the research costs alone for each drug candidate ready to enter clinical development are in the order of US$20 million. So the drug-discovery costs for a minimum of 48 clinical candidates that could result in 8 new drugs are probably close to US$1 billion — more than the annual research budget of most drug companies.

A new collaborative mindset is required if we are to scale up drug discovery for tropical diseases. The TDR is now implementing a scalable, collaborative model for drug discovery that is based on networks and partnerships with industry and academia in developed and developing countries[3]. Scaling up or expanding innovative drug-discovery networks could be a way to mobilize resources from both public and private sectors. This approach would involve the creation of a transparent and flexible global portfolio of drug-discovery projects and research requirements. Capacity for drug discovery worldwide would then expand through the pooling of private- and public-sector resources, and the productivity of existing non-profit research would increase through the sharing of information, tools and ideas.

From lab to clinic

Investment in basic research for tropical diseases has created the scientific opportunity to develop a new generation of drugs. Compared with therapies for complex diseases, such as heart disease, type 2 diabetes and cancer, those for tropical diseases are generally simpler to test and optimize. This is because it is often possible to test compounds using highly predictive *in vitro* or *in vivo* screens in which death or elimination of the infectious organism can be evaluated directly. If a candidate successfully kills a parasite in a dish and a laboratory animal model, then there is a good chance that it will be effective in humans when it moves to clinical development, although further safety and dosing criteria will still need to be met.

WHO/TDR/LSTM

But many organisms that cause tropical diseases are difficult to maintain in the laboratory because of their complex life cycles. So it makes sense for specialist facilities and disease experts to work in a coordinated manner as screening centres for multiple drug-discovery groups, as in the TDR Screening Network.

New targets for drugs are also being identified as researchers finish sequencing the genomes of pathogenic organisms. Compared with whole-organism screening, mechanistic targets such as purified enzymes or receptors are usually more suitable for rapid screening using the libraries of many hundreds of thousands of compounds that are available in some pharmaceutical companies. Valuable lead compounds can be missed in whole-organism assays when they are rapidly metabolized or are unable to cross cell membranes to reach their mechanistic target. Unfortunately, genomics-based approaches have so far failed to deliver the predicted new generation of anti-infective drugs. This stems in part from a focus on the biological relevance of potential targets rather than on their 'druggability'[4] (the

Few effective medicines are available to treat diseases prevalent in tropical regions.

probability that the targets will bind to drug-like molecules).

Chemogenomics is therefore emerging as a useful strategy for target-based drug design[5] (see 'Drug discovery: from the genome to the clinic', overleaf). This approach exploits both chemical (for example, protein structure–activity data and drug binding-site features) and genomic (DNA sequence) information about the pathogenic organism. Its power lies in the ability to relate targets in a parasite back to likely chemical starting points *in silico*, enabling potential targets to be selected before expensive and time-consuming drug-screening and optimization studies are undertaken[6]. Researchers can assign priority to the most promising mechanistic drug targets for further investigation, building up a portfolio of potential projects, which can then be ranked on the basis of 'drug-hunting' criteria. Identifying promising chemical leads as early as possible could lower failure rates from screening initiatives and help reduce overall discovery costs.

Global portfolio

As well as being clinically effective, pharmaceuticals for developing countries need to be cheap to manufacture, stable during distribution and storage, and easy to administer to ensure wide usage. About half of the drugs being developed to treat neglected diseases fail some of these criteria[7], reflecting a lack of attention given to optimizing for desired properties, a paucity of lead structures and a reluctance to abandon the few unpromising leads. An agreed set of product profiles that describe the target clinical efficacy, route of administration, dosage regimen, safety and cost of treatment are widely used in industry to drive lead discovery and candidate selection. Emphasis should be placed on understanding the product profile of drugs required for the various neglected diseases[3].

One powerful way to help align global research on infectious diseases with patient needs is to use challenge-based innovation methods. Communication of specific public challenges by appropriate bodies could act as an impetus for innovation (see page 164).

Moreover, the efficiency of drug discovery would improve if a shared public portfolio of prioritized drug targets and candidate drug structures for optimization were available, because this would stimulate the pooling of research resources worldwide and attract new contributors. The recently announced TDR Drug Target Prioritization Database (http://tdrtargets.org) provides a central facility for information on drug targets, with targets ranked according to their predicted druggability, potential importance to the organism and selectivity compared with related human proteins. Specific risks and opportunities can be assessed from the known and inferred molecular properties of each putative drug target.

By listing specific projects and requirements, these portfolios could help to mobilize new resources. Interested participants could identify collaborators, and this would increase the diversity of expertise in, for example, optimization of lead candidates, screen development or drug screening. But each organization would carry out only modular research tasks appropriate for its facilities and expertise.

Innovation networks

Scaling up drug-discovery capacity for neglected diseases means designing a mechanism that is attractive to all stakeholders — such as industry, academia, governments and international agencies — involved in drug discovery. Despite a lack of market incentive, about half of the research projects currently focusing on neglected diseases are conducted by pharmaceutical companies. Several large companies, most notably GlaxoSmithKline, Novartis, AstraZeneca and Eli Lilly, have founded research institutes dedicated to research into tropical diseases. Industry also makes a substantial contribution in kind to public–private product-development partnerships. Also, in recent years, funding from governments and philanthropic foundations has helped to establish new drug-discovery units for tropical diseases in developed and developing countries.

More industrial enterprises and academic groups could be persuaded to participate if they could contribute advice, skills and infrastructure while avoiding the sustained and costly overheads of running separate, dedicated institutes or programmes, and if they could safeguard intellectual property through appropriate contractual arrangements. 'Virtual' drug-discovery networks may be the mechanism to enable research into neglected infectious diseases to be scaled up markedly, by attracting new private and public participants.

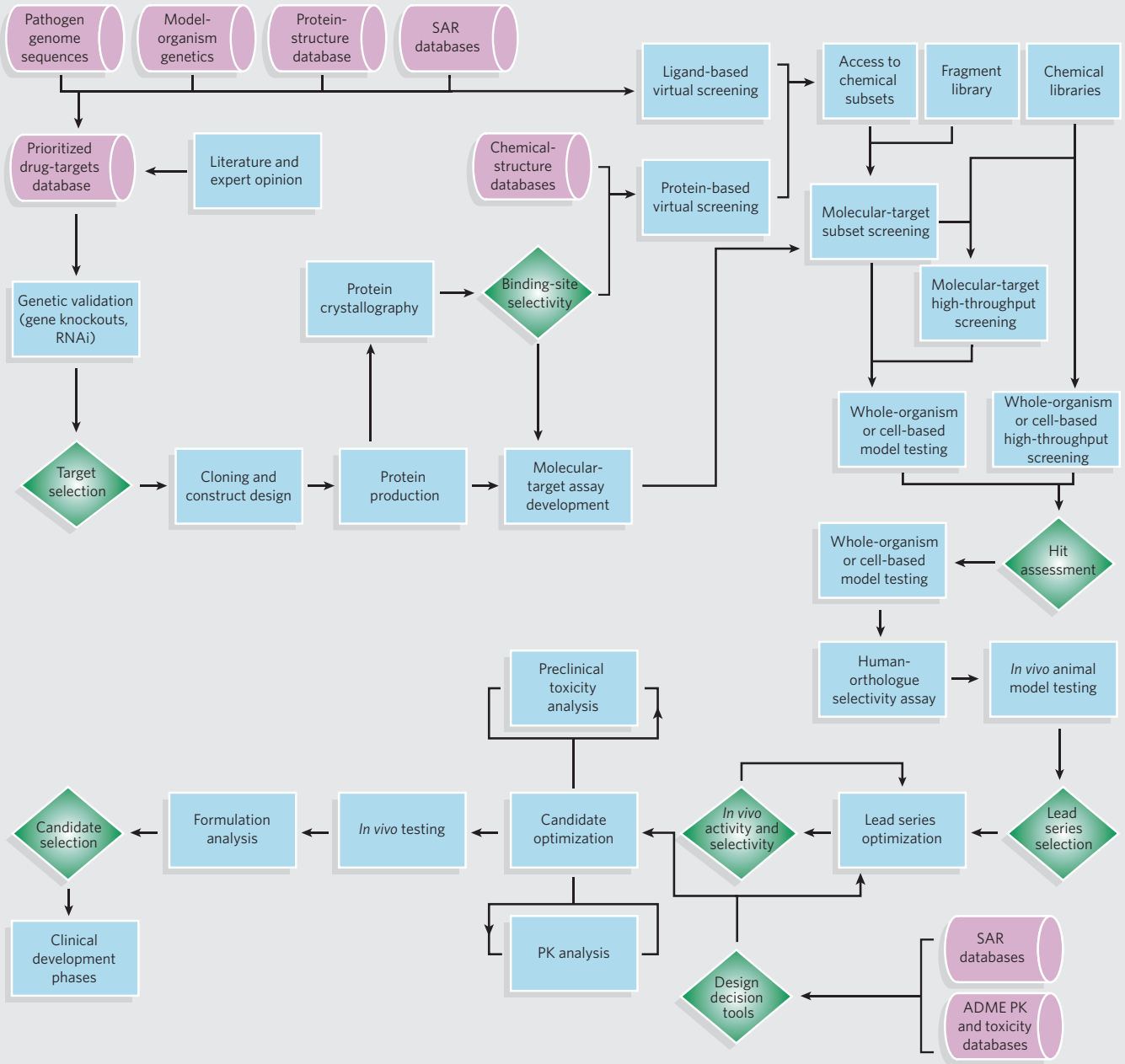
Partnering industry

Partnerships between the pharmaceutical industry, the public sector and non-governmental organizations are the foundations for the success of the innovation-network concept. By creating a framework in which industry can contribute to its expertise, knowledge, compound libraries and infrastructure, considerable cost savings could be obtained.

Industrial ventures might contribute expertise with sponsored sabbatical or fellowship schemes, or training in medicinal chemistry, for example, which is rare in the public sector. Industrial scientists could form part of a virtual multidisciplinary project team, collaborating with scientists at public and private institutes so that skills and experience are transferred in

DRUG DISCOVERY: FROM THE GENOME TO THE CLINIC

Drug discovery consists of discrete decision stages (diamonds), with iterative experimentation at each step (boxes). The early stage involves identifying 'hits' by exploiting information (cylinders) from the biochemical properties and genome sequence of an organism in a lead-discovery strategy. Proteins that are found to be essential by chemogenomic analysis and are predicted to have promising drug-like properties can be prioritized for genetic validation. Suitable molecular-screening assays for validated targets are subset screening (10–1,000 compounds), fragment screening (1,000–2,000 compounds) and high-throughput screening (10,000–1,000,000 compounds). Chemoinformatics and 'virtual' screening methods can be used on selected subsets of compounds from various databases and then tested before scaling up an assay for high-throughput screening. Hits that pass this step can then be assessed for efficacy, selectivity and druggability. Successful hit compounds are then usually clustered into chemical series, which need to be chemically optimized into lead compounds with suitable efficacy and selectivity. Lead compounds are then optimized for their pharmaceutical properties (absorption, distribution, metabolism and excretion; ADME) and absence of overt drug toxicity. This is an iterative, time-consuming process of medicinal-chemistry design, which is assisted by design-decision tools. Ideally, several lead series are run in parallel to increase the overall chance of success. PK, pharmacokinetics; RNAi, RNA interference; SAR, structure–activity relationships. **A.L.H., M.J.W. & S.N.**




PFIZER

Pharmaceutical companies, such as Pfizer (pictured), are increasingly sharing tools with non-profit organizations to accelerate drug discovery.

both directions, helping to build global innovation capacity.

The infrastructure of industrial laboratories for research into tropical diseases could be made accessible to sponsored scientists within a collaborative framework. This might include access to facilities such as high-throughput screening equipment, automated synthesis and computer-aided drug design, under joint academic–industrial doctoral and postdoctoral training programmes. The TDR postdoctoral fellowship programme, which funds scientists to work in pharmaceutical companies on specific tropical-disease projects, is an example of how public funding can be used in an effective collaboration with industry.

Pharmaceutical companies can also provide access to a subset of their extensive libraries of proprietary chemical compounds — on a contractual and usually confidential basis — to help spread their corporate mandate across a broad range of tropical diseases. The TDR has ongoing partnerships with Pfizer, Merck Serono and Chemtura that exemplify this approach: subsets of compounds are selected by companies for screening by the TDR Compound Evaluation Network (which follows the process shown in the middle portion of the graphic, opposite).

Because shipping large amounts of proprietary chemicals to external screening laboratories can be costly and impractical, several high-throughput screening campaigns are ongoing within companies against tropical-disease targets. Virtual screening of patent and literature compound databases could also be used to help companies and collaborators select small subsets of compounds for initial screening experiments.

> "All of the tools needed to create a sustainable and scalable model of drug innovation are within our reach."

Making global networks work

Virtual drug-discovery networks[8] for infectious diseases prevalent in the developing world are already functioning on a small scale. Various researchers and institutes are called on by public–private partnerships to contribute their diverse expertise, technology and resources for a limited duration. Compared with established pharmaceutical companies, which have a relatively fixed resource and skills base, these networks have the flexibility to reprioritize projects and to outsource as needed to academia, industry and contract research organizations.

The TDR has developed a virtual drug-discovery capability by using a series of portfolio, screening, medicinal chemistry and ADME (absorption, distribution, metabolism and excretion) networks. But to be scalable and sustainable, additional components, including funding, are required to attract and coordinate more participants from the private and public sectors. To work effectively, innovation networks need to be coordinated. This includes communicating the portfolio to network participants and beyond; establishing win–win collaborations; and encouraging innovative measures to lower the costs of drug discovery and delivery. Within the framework of an enhanced innovation network, large and small companies can make worthwhile contributions to the overall mission depending on what level of commitment they can afford. This concept is already being implemented by the TDR on a small scale.

Pharmaceutical companies contain business units that benefit from shared services and enabling technologies to support their portfolio of projects. Likewise, network 'enablers', such as shared contract services, discovery tools, databases and knowledge management, would help the public–private product-development partnerships and drug-discovery networks to scale up their activity.

A coordinated, information- and knowledge-sharing 'clearing house' would be a key requirement for the active management of an enhanced virtual drug-discovery network. It would provide a place to attract new participants to public–private drug-discovery partnerships and match them to individual portfolio projects and collaborators, while brokering confidentiality agreements and contract research[9]. The modular nature of the network would allow projects to be undertaken by self-organizing, self-motivated, virtual teams with a common goal. Productivity would be increased by sharing, for example, ADME and toxicology data, as well as workflows and research tools, such as physicochemical prediction models.

Our capacity to combat neglected tropical diseases must now be mobilized to include a pan-industry effort using an open innovation[10] approach to drug discovery. All of the tools needed to create a sustainable and scalable model of drug innovation for many of these devastating scourges are within our reach. The challenge now is for industry, governments and philanthropists to unite in undertaking this mission. ■

Andrew L. Hopkins is head of chemical genomics at Pfizer Global Research and Development, Sandwich, UK. Michael J. Witty is group director of global portfolio management at Pfizer Animal Health, Sandwich, UK. Solomon Nwaka is leader of drug-discovery activities at the Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland.

1. Pink, R., Hudson, A., Mouries, M.-A. & Bendig, M. *Nature Rev. Drug Discov.* **4,** 727–739 (2005).
2. Kola, I. & Landis, J. *Nature Rev. Drug Discov.* **3,** 711–715 (2004).
3. Nwaka, S. & Hudson, A. *Nature Rev. Drug Discov.* **5,** 941–955 (2006).
4. Payne, D. J., Gwynn, M. N., Holmes, D. J. & Pompliano, D. L. *Nature Rev. Drug Discov.* **6,** 29–40 (2007).
5. Bredel, M. & Jacoby, E. *Nature Rev. Genet.* **5,** 262–275 (2004).
6. Hopkins, A. L. & Groom, C. R. *Nature Rev. Drug Discov.* **1,** 727–730 (2002).
7. Moran, M., Ropars, A.-L., Guzman, J., Diaz, J. & Garrison, C. *The New Landscape of Neglected Disease Drug Development* (London School of Economics/Wellcome Trust, London, 2005).
8. Nwaka, S. & Ridley, R. G. *Nature Rev. Drug Discov.* **2,** 919–928 (2003).
9. Munos, B. *Nature Rev. Drug Discov.* **5,** 723–729 (2006).
10. Chesbrough, H. *Open Business Models: How to Thrive in the New Innovation Landscape* (Harvard Business School Press, Boston, 2006).

Acknowledgements We thank G. Samuels, T. Wood, B. Callan and R. Ridley for useful discussions.

A prescription for drug delivery

Improvements in basic infrastructure are the key to saving millions of lives each year, say **Julian Lob-Levyt** and his colleagues.

One-quarter of all child deaths result from diseases that could be prevented by vaccination. But the introduction of new and underused vaccines into the world's poorest countries is hampered by inadequate infrastructure and a lack of guaranteed finance.

In wealthy countries, it is taken for granted that research and development of new vaccines and other medicines is the first step towards the manufacture of effective products and their distribution through carefully regulated channels. But this process breaks down in the poorest countries. There is a lack of investment in research and development in these regions because the pharmaceutical industry perceives that the market for new drugs is limited or non-existent. Also, the basic infrastructure required for distribution — from human resources to diagnostic tools — is badly in need of strengthening.

> "The health gains made in Europe in the past 150 years could be achieved in Africa within the next 10 to 20 years."

If a country is to capitalize on the benefits of new advances in the fight against infectious disease, then it needs to deliver basic services to all those in need, including those living in hard-to-reach areas. The ability to do this depends on tight budgets and the relative affordability of products and services: tough choices must be made in the allocation of scarce resources.

Ethiopia, one of Africa's poorest nations, is an example of a country where the weak infrastructure for health-care delivery is a major impediment to providing basic medical services. As Ethiopia's health minister, Tedros Ghebreyesus, has said, "Our vehicle has not been strong enough to carry all the programmes we have loaded on it. Now we are working to strengthen the vehicle so that it can carry our programmes, the vaccines and the other health-care interventions, to every corner of this vast country."

To address the challenges posed by weak infrastructure and to accelerate access to new and underused vaccines in about 70 of the world's poorest countries, including Ethiopia, the GAVI Alliance was launched in 2000. GAVI brings together the main public- and private-sector stakeholders in vaccination, including national governments, philanthropic organizations, the vaccine industry and international bodies.

The alliance provides financial support for vaccination against hepatitis B and *Haemophilus influenzae* type b (Hib) to countries that show more than 50% coverage with the vaccine DTP3 (used worldwide to protect against diphtheria, tetanus and whooping cough). Countries falling short of this threshold are not eligible for funding to roll out these additional vaccines but are instead eligible to apply for support to strengthen their current vaccination schemes. GAVI's financial support is tied to indicators of performance, and success is rewarded.

Unlike the traditional approach to financing health systems, in which priorities are set externally and/or tied to the purchase of specified products and services, GAVI provides support that is not earmarked. This allows countries to assign their own priorities to funds. In general, countries supported by this scheme focus on training, management and improving infrastructure.

This type of funding has resulted in swift changes. Fifteen million additional children have been vaccinated in the seven years since GAVI's launch. In eligible countries, the overall coverage of vaccination with DTP3 increased from 63% in 1999 to 71% in 2005 (see graphic, right). The figures are particularly impressive in the African region — DTP3 coverage increased from 44% in 1999 to 65% in 2005.

Much of this increase in DTP3 coverage can be attributed to the immunization services support provided by GAVI to improve vaccine delivery (C. Lu *et al. Lancet* **68,** 1088–1095; 2006).

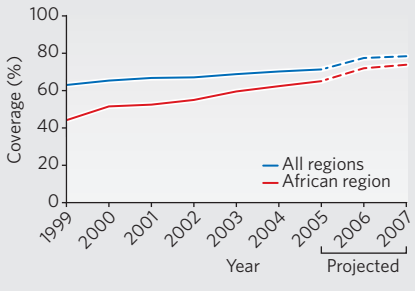
The African-led success story defies common misperceptions of poor performance in Africa. Marked increases in the uptake of vaccines against hepatitis B and, more recently, Hib have also occurred (see graphic, opposite).

C. NESBITT/GAVI

Weaknesses in the vehicle

With results such as these, the potential impact of strengthening infrastructure on a much larger scale is evident. Under similar circumstances, the poorest countries could probably effect similar changes in other medical areas — making the health-related Millennium Development Goals, set by the United Nations, potentially achievable. But, despite an increase in financial commitments as donors recognize the large sums required, the

DTP3 VACCINE COVERAGE IN COUNTRIES ELIGIBLE FOR GAVI SUPPORT



GAVI ALLIANCE PROGRESS REPORT 2006

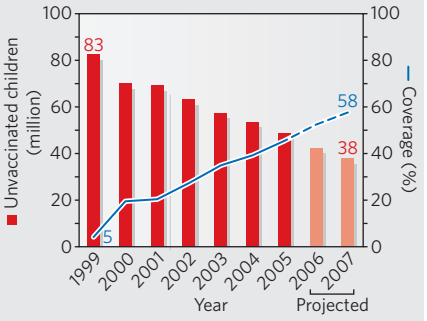
Better coverage: Nigerian women receive free bednets when their children have been fully vaccinated.

disbursement of funds to the poorest countries still needs to be scaled up, and there is a tendency among donors to focus on disease-specific issues at the expense of broader reinforcement of health systems. This means that long-term planning by health ministers in the poorest countries to 'strengthen the vehicle' remains a major challenge.

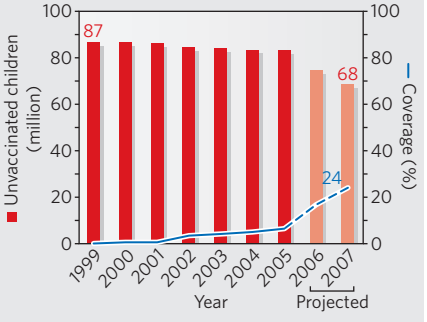
Although increasing funds are flowing through partnerships such as GAVI and the Global Fund to Fight AIDS, Tuberculosis and Malaria, these organizations still face crucial constraints. An important issue is the acute shortage of staff, particularly in sub-Saharan Africa. Even where staff are available, they can lack appropriate skills, or their skills are in demand elsewhere. So recruitment, training and retention pose a challenge.

There are also shortages of drugs and vaccines where they are needed. Infrastructure, logistics and management systems are inadequate to meet the scale of need and might not even extend to the communities they are meant to serve. Because of cultural or financial reasons or poor quality, the services might not meet demand or indeed be accessible or accountable to the poor and marginalized (see page 160). Governments often rely too much on delivery by the public sector.

HEPATITIS B VACCINE COVERAGE IN GAVI-ELIGIBLE COUNTRIES



HIB VACCINE COVERAGE IN GAVI-ELIGIBLE COUNTRIES



Such constraints are not new. But the considerable increase in funding through global health partnerships, the high profile of these partnerships and their need to be accountable in terms of results means that these constraints must be confronted anew and that more radical solutions must be considered. The international community has been successful in mobilizing financial and political support for HIV/AIDS prevention and treatment, and for vaccines for children. It has been less effective, however, at uniting around a vision and strategy to tackle the more fundamental challenge — the building of an integrated delivery platform across the public, private and civil-society sectors in the poorest parts of the world.

A mother cannot afford to go six times in a week, often to different places, to access separate services to immunize her child, treat her son's asthma, ensure she has antiretroviral drugs for her husband, collect an insecticide-treated bednet, receive family-planning advice and discuss whether her daughter should receive a vaccine against cervical cancer. Yet this scenario will remain the reality for millions of women in the absence of rational and functional health-care provision.

Financing the vehicle

GAVI has responded by creating a stream of finance so that health systems can overcome such practical difficulties. Alone, this finance is not enough. But if it is used in conjunction with other donations, and if countries form partnerships around a joint, agreed strategy, then financial donors might change their behaviour. Frankly, it is not complicated. Capacity building and additional funds will be required. Above all, political leadership — both nationally and globally — and agreement on a coordinated and sustained effort are vital. The goal is to deliver better health to women, children and other vulnerable groups.

GAVI has developed other mechanisms to overcome some of the present limitations of direct bilateral aid. The International Finance Facility for Immunisation (IFFIm), launched

GAVI ALLIANCE PROGRESS REPORT 2006

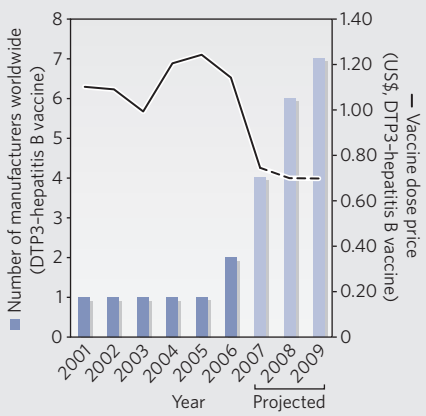
E. PURNOMO/GAVI

If it weren't for infrastructure improvements in Indonesia, this baby might have missed out on a life-saving vaccine.

in 2006 by GAVI and supported by the governments of Brazil, France, Italy, Norway, South Africa, Spain, Sweden and the United Kingdom, has led the way in doing development business differently. IFFIm takes long-term (20-year), legally binding commitments from donors and borrows against them in the capital markets, producing upfront finance. This means that governments can make long-term plans for strengthening their health systems based on a predictable flow of funds. An anticipated IFFIm investment of US$4 billion is expected to prevent 5 million child deaths between 2006 and 2015, and more than 5 million adult deaths from liver disease associated with hepatitis B. The IFFIm is a win–win model — a financial instrument that saves lives while producing a high return on investment.

Other innovative financial measures are necessary to bridge the breakdown in the road from development to delivery. For example, vaccines that would prevent millions of deaths face long delays before they are developed, tested and produced for use in the poorest developing countries. Advance market commitments (AMCs) have recently been launched to tackle this market failure. These are financial commitments to subsidize the future purchase of a vaccine, up to a pre-agreed price, provided that an appropriate vaccine is developed and that there is still a demand when it is produced. By guaranteeing that the funds will be available to purchase vaccines once they are developed and produced, the AMC mimics a secure vaccine market and takes away the risk that countries might not be able to afford a high-priority vaccine (see page 176). In addition, vaccine prices decline as demand is generated and new manufacturers enter the market (see graphic, left).

DECLINE IN VACCINE PRICES IN DEVELOPING COUNTRIES



GAVI ALLIANCE PROGRESS REPORT 2006

In the long term, the considerable size and growth of GAVI and the prices that it has negotiated, together with the introduction of specific market-shaping mechanisms such as AMCs, should partly address the market failure that keeps new health-related technology from the poor people who need it.

The next level

Today, the broader challenge is still to make vaccination programmes sustainable. Technologies such as vaccines and antiretroviral drugs have the potential to deliver a generational leap in achieving the Millennium Development Goals. The health gains made in Europe in the past 150 years could be achieved in Africa within the next 10 to 20 years. But without an accelerated and coordinated effort to tackle the fundamental constraint — weak infrastructure for the delivery of basic health care — the full potential of innovative strategies will not be realized.

> "Without an accelerated and coordinated effort to tackle weak infrastructure, the full potential of innovative strategies will not be realized."

In the next decade, vaccines against diseases such as pneumonia, malaria and tuberculosis, as well as infection with rotavirus or human papilloma virus, should be rolled out successfully. Ensuring that the new vaccines make it to remote villages in deprived areas must be at the centre of global health efforts. The alternative — millions more unnecessary deaths and millions more dollars lost — is unacceptable.

For the global health community, the challenge is therefore to work collectively to secure long-term political and financial support so that poor countries can develop integrated health-service delivery platforms, which are fundamental to productive nations. ■

Rebecca Affolder is head of the executive office of the GAVI Alliance, Geneva, Switzerland. Ivone Rizzo, Craig Burgess and Abdallah Bchir are senior programme officers at the GAVI Alliance. Julian Lob-Levyt is executive secretary of the GAVI Alliance, and chief executive officer and president of the GAVI fund.

Patent sense

Protecting intellectual property saves lives in the developing world, argues **Paul Herrling**.

Non-profit institutes funded by industry carry out R&D on drugs for neglected diseases.

Many diseases are endemic in the developing world, yet for a number of these there are few safe and effective treatments. This lack of medicines results from an industrial model that has been in place for more than 50 years. Basic scientific research carried out in the public sector is translated into life-saving medicines mainly by pharmaceutical companies. This is a lengthy, onerous and expensive process — taking about 15 years and costing hundreds of millions of dollars per drug — and comes with a high risk of failure. Nevertheless, more than 90% of new molecular entities discovered and developed as medicines between 1990 and 1999 originated from pharmaceutical companies[1,2].

Drug firms may be the main source of new therapies, but they remain commercial entities that can invest the considerable resources required to translate basic science into an effective medication only when there is a reasonable chance of financial return.

There is little opportunity to get an adequate return on investment for infectious diseases such as tuberculosis (TB), dengue fever, malaria, leishmaniasis and African trypanosomiasis (sleeping sickness), which mainly affect people living in resource-poor regions. In other words, market mechanisms fail in these cases, and there is insufficient drug-discovery research and development (R&D) for these common infectious diseases.

> "In the absence of a patent, the only way inventors can protect their inventions is through total secrecy."

No secrets

Some organizations interested in improving access to medicines in the developing world, such as Médecins Sans Frontières and Oxfam, think that a major impediment to affordable medicines is the patent system. But this is not the case. This system protects intellectual property in countries whose economies are based, to a large extent, on innovation. A patent is defined as a grant by the state of exclusive rights for a limited time in respect of a new and useful invention. These rights usually imply that, for a limited time, only the innovator, or a person or entity licensed by the innovator, can sell products based on the invention. This offers the innovator an opportunity to recover the investment needed to develop the invention into a practical product. Without this incentive, important discoveries would never be developed into useful products. Modern patent law provides protection for 20–25 years, which should be compared with the 15 years, on average, needed for the discovery and development of a new drug. In return for these rights, the innovator discloses a description of the invention that allows other experts to reproduce the key findings. This process is firmly based on the premise that knowledge is gained only through a full understanding and appreciation of previously published advances.

In the absence of a patent, the only way inventors can protect their inventions is throughtotal secrecy, which is counter to furthering innovation, a fact often ignored by those who consider that patents prevent research. It is only when patents are used excessively to protect information — to the extent that researchers cannot use a patent-protected invention in their studies — that the system is a considerable barrier to further innovation. To prevent such abuse of patents, several countries have implemented the 'research exemption', which allows scientists to use patent-protected technology freely for their research provided they do not exploit it commercially. In light of these issues, the protection of intellectual property with patents is crucial for pharmaceutical companies to discover and develop new drugs for the developing world.

Neglected no longer

In the past decade or so, the drug industry has formed partnerships with the public sector, generating pipelines of early-stage potential medicines for certain neglected diseases. These partnerships include the Global Alliance for TB Drug Development, the Drugs for Neglected Diseases Initiative (DNDi) and the Medicines for Malaria Venture. In 2004, 63 new drugs were being pursued by this approach[3].

Most of these drugs originate from R&D programmes in pharmaceutical companies and are patented accordingly. But these patents are not used to enforce unaffordable prices in the developing world or to prevent manufacturers from selling generic versions of these drugs in developing regions.

An example of this is Coartem, an artemisinin-based combination therapy. One of the most effective treatments for malaria

NOVARTIS

at present[4], Coartem is included on the list of essential medicines by the World Health Organization[5]. One of its components was discovered by Chinese scientists, and this was then clinically characterized, and developed and produced as a combination therapy, by Novartis. Coartem is patented in 49 countries but is available at cost price to patients in all countries in which malaria is endemic. In 2006, Novartis delivered more than 62 million treatments of Coartem to more than 30 countries at cost price. A generic version of the drug is being manufactured by Indian pharmaceutical companies and sold in the developing world.

Another example is a new artemisinin-based combination therapy for malaria, developed by sanofi-aventis in collaboration with the DNDi. This treatment, also on the World Health Organization's list of essential medicines[5], will be available in most of sub-Saharan Africa next year. Sanofi-aventis has set aside its patent rights and will supply the medicine at cost price to the poorest populations in countries in which malaria is endemic.

In general, drugs developed by pharmaceutical companies as part of public–private partnerships are patented by industry, as it is the discoverer. But they are exclusively licensed for treating neglected diseases in agreements stipulating that the drugs will be available at cost price to the developing world.

Necessary system

So why are patents even necessary? There are three main reasons. First, commercial entities, including drug companies, can allocate resources to non-profit projects only if they are financially sound. Research- and innovation-based companies earn sufficient returns on their R&D investments only if they are permitted a marketing-exclusivity period — accorded by patents — for their innovative products. In countries without this protection in place, the drug industry is not research intensive and innovation based. Indeed, this was the case in India while there was no patent protection for new pharmacologically active molecules.

The randomness of evolution provides a second reason for patenting drugs for neglected diseases. Nature does not distinguish between the biology of diseases of the poor and the rich. So therapeutic molecules or pathways that are targeted by drugs for neglected diseases might also be relevant for treating diseases that affect people in more affluent regions. For example, the Novartis Institute for Tropical Diseases in Singapore takes all molecules that have activity against dengue virus and systematically tests them against the West Nile and hepatitis C viruses, which belong to the same family as dengue and cause disease also in developed countries. If a molecule has activity against dengue virus or another family member, it is developed and patented. Considerable financial returns could be generated in the developed world, and a portion of these is earmarked for refinancing, and for providing long-term sustainability to, the institute's non-profit initiatives.

A third reason to patent such drugs relates to emerging economies such as those of Brazil (see page 180) and India. In these countries, there are populations of very poor patients, and the non-profit model would certainly apply. But in the same countries, there are growing numbers of more affluent patients, who are increasingly able to buy their medication, either directly or through health insurance. For these patients, the company that developed the medicine could expect to generate revenue as a result of patent protection. Such differential pricing within a country (see page 176) would encourage further innovation not only by global companies but also by local enterprises. This model would require legislation that prohibits copying before patents have expired but that allows generic-drug production after patent expiry.

For the foreseeable future, the discovery and development of new medicines will be driven almost exclusively by commercial pharmaceutical companies. The only way that these firms can remain viable is through a robust intellectual-property protection system. This system, therefore, contributes to saving lives. Without it, there would be few new drugs for any disease, regardless of whether it afflicts the rich or the poor. ■

Paul Herrling is head of Corporate Research, Novartis International, Basel, Switzerland.

K. PRINSLOO/AP

Aid organizations such as Médecins Sans Frontières say that patents block access to medicines.

1. Reichert, J. M. & Milne, C. P. *Am. J. Ther.* **9,** 543–555 (2002).
2. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. *J. Health Econ.* **22,** 151–185 (2003).
3. Moran, M. *PLoS Med.* **2,** e302 (2005).
4. Mutabingwa, T. K. *et al. Lancet* **365,** 1474–1480 (2005).
5. World Health Organization Model List of Essential Medicines. www.who.int/medicines/publications/EML15.pdf (2007).

At what price?

Differential pricing could make global medicines affordable in developing countries. But drugs for diseases that have no market in the developed world will require additional subsidies, says **Patricia M. Danzon**.

P. PARKS/AFP/GETTY IMAGES

For the general population in developing nations to have appropriate access to medicines, existing drugs must be affordable, and innovation is needed to develop new medicines. But this presents a potential conundrum: prices that are high enough to pay for research and development (R&D) may make medicines unaffordable in developing regions. Differential pricing[1] (also known as price discrimination) can offer a solution to this dilemma, at least for drugs with considerable sales in the developed world. Prices in affluent countries — and to a lesser extent in middle-income countries — could generate sufficient revenue to pay for R&D, whereas prices in developing nations need only cover their marginal costs. But differential pricing will be possible only if market separation can be sustained, preventing the low prices in developing countries from spilling over to higher-income nations. However, for drugs that treat diseases endemic only in the developing world, sales are insignificant in the developed world, and additional subsidies are essential to attract R&D for these diseases.

> "Today, R&D costs about US$1 billion for each new drug."

Economics of differential pricing

This prescription of differential pricing and separate markets for on-patent pharmaceuticals is at odds with the free-trade and global-pricing maxims generally favoured by economists. The reason is that R&D costs roughly US$1 billion for each new drug approved in 2007, including the cost of failures and the necessary return on capital invested over the 8–12 years required for R&D[2]. Pharmaceutical R&D can benefit patients globally, raising the question of how this joint cost should be allocated among consumers to generate the greatest benefit. Counting all consumers equally, the answer is that prices should vary inversely with the consumers' price sensitivity — a theory of optimal differential pricing known as Ramsey pricing[3]. Prices must exceed marginal production cost for at least some users in order to pay for R&D. But if all consumers are charged the same price, then the most price sensitive will reduce their use and lose more benefit than would the less price-sensitive consumers. In practice, such sensitivity to drug prices is hard to measure, but a reasonable assumption is that it varies inversely with income. Ideally then, countries with lower per-capita income should be charged less than countries with higher per-capita income.

More generally, economic theory shows that differential pricing promotes greater social welfare than uniform pricing if consumers in aggregate buy more under differential pricing[4] — which seems plausible for pharmaceuticals. A simulation[5] comparing worldwide pharmaceutical prices, revenues and number of consumers served under a single global price with differential pricing between national markets (that is, one price per country) found that differential pricing increases consumer access to drugs by a factor of roughly 4–7 compared with uniform pricing. In addition, differential pricing within, as well as between, countries could significantly increase affordability for poor populations in countries that have a skewed income distribution and no national health insurance. Differential pricing would not only increase the use of existing drugs (static efficiency) but should also increase R&D and the flow of new drugs as a result of increased sales revenue (dynamic efficiency)[6].

A common objection to differential pricing is that it 'shifts costs' between low- and high-priced markets (see ref. 7, for example). But this argument implicitly assumes that the joint costs of R&D should be allocated equally to all users and/or that manufacturers engage in cost-plus pricing, such that if some consumers pay less, others automatically pay more. But, if markets

Out of reach: many medicines are too expensive for people in developing regions.

are separate, manufacturers set the price for each market based on local conditions, irrespective of prices elsewhere. Thus prices would not automatically fall in high-priced markets such as the United States if low-priced markets were to pay more. It is true that if manufacturers were required to charge a uniform price worldwide, prices might drop in the United States because the single price would be based on a weighted average of price elasticities in all the major markets. However, this could be viewed as 'free riding' of high-income, price-insensitive countries on price-sensitive, lower-income countries, and not as the elimination of cost-shifting.

Implementing differential pricing

Although there is widespread support for differential pricing of medicines used to treat HIV/AIDS, tuberculosis and malaria in the lowest-income countries, there is no consensus on applying differential pricing more generally to other drugs and to middle-income countries, or on appropriate benchmark prices and differentials for different countries. Setting benchmark prices based on costs is unworkable because accounting costs do not capture all relevant R&D costs — including failures and the necessary return on funds invested. Moreover, setting prices based on costs creates perverse incentives for producers to inflate costs. More generally, achieving differential pricing through regulation would be vulnerable to political pressures, because underpricing immediately benefits current consumers, but its negative effect on the flow of new drugs is not evident for 8–12 years and will be hard to attribute to specific policies or politicians.

Fortunately, a regulatory structure is not needed to achieve appropriate price differentials. If markets are separate and reasonably competitive, the price differentials that manufacturers would voluntarily charge to maximize profits are similar to the Ramsey optimal price (ROP) differentials required to maximize welfare. Absolute prices may differ, however, because ROP prices are intended to cover costs with a normal return on capital, whereas actual prices can yield positive profits or fail to cover costs, depending on market conditions, competition, regulation and other factors.

Obstacles to differential pricing

Although manufacturers have incentives to pursue differential pricing that is roughly related to per-capita income, the limited evidence indicates that prices are relatively high compared with income in poor countries, making many drugs unaffordable[8,9]. Several factors undermine differential pricing in practice.

Manufacturers adopt differential pricing only if markets remain separate, such that low prices in one market do not erode potentially higher prices in other markets. In fact, such price spillovers occur increasingly, as a result of parallel trade (drug importation by intermediaries, to profit from price differences) and because regulators in middle- and high-income countries use lower foreign prices as an 'external' reference to cap their own domestic prices.

Under traditional patent rules, a patent holder can bar unauthorized importation of a product. However, the European Union has legalized parallel trade between its member states, on grounds of free trade. In the United States, legislation authorizing parallel importation has been enacted but not yet implemented because requirements for quality assurance and cost savings have so far not been met[10]. Although proponents argue that parallel trade is just free trade, in fact parallel trade in pharmaceuticals usually results from differences in price regulation or in per-capita income, and not from lower, real resource costs. Parallel trade, therefore, offers none of the usual efficiency gains from trade; it may, in fact, increase resource costs due to transportation, quality control and relabelling, and reduce welfare gains that would result from differential pricing. Most of the savings accrue to the intermediaries and not to the consumers or payers in the importing country, who continue to pay the higher price.

External referencing can have an even greater impact than parallel trade. It is formally incorporated into drug-price regulation in many countries, including Canada, Greece, Italy, Japan and the Netherlands, and is used informally by many others — including the United States and the United Kingdom, where comparison of international prices frequently informs and influences drug policy. In addition, Brazil has recently demanded the lowest price granted to any other purchasers. With external referencing, manufacturers are reluctant to price a drug cheaply in one country if this would undermine potentially higher prices in other countries. Companies often try to keep the launch price of a drug within a narrow band, preferring to delay or not launch in countries that do not meet the price target[11]. Although

I. BERRY/MAGNUM PHOTOS

Differential pricing, coupled with subsidies, could be the key to getting medicines to those who need them.

some middle- and high-income countries may benefit in the short run from external referencing, in the longer term these countries will also be worse off, as the breakdown of differential pricing reduces drug sales and hence leads to less R&D and fewer new medicines.

The distribution of income in many low- and middle-income countries is skewed, causing manufacturers to aim medicine prices at the small, high-income subgroup. Some companies offer discounts to public clinics and other programmes that target the poor, but most governments are unwilling to accept differential pricing within countries. Even where public clinics in principle offer free drugs, many poor people buy medicines from private pharmacies because the clinics are far away, entail long waits or do not have the drugs[12].

Retail drug prices include distribution margins for the wholesaler and pharmacy, which are often higher in poor countries than in higher-income countries that have competitive and technically sophisticated distribution systems, and/or powerful payers to negotiate discounted dispensing fees. Many developing countries impose high import tariffs, inserting an additional wedge between the price paid by the consumer and the price charged by the manufacturer.

Maintaining market separation

Achieving sustainable differential pricing requires market separation, which in turn requires policies and institutions that prevent price spillovers from low-income countries to middle- and high-income countries.

Although the World Trade Organization permits countries to authorize parallel trade, middle- and high-income countries should adopt patent and other policies to bar drug importation, particularly from lower-income countries. It is particularly important that the United States maintains its bar on drug importation. Even if importation is authorized only from high-income countries, enforcement will be unable to prevent importation from other countries if the price differentials offer large profit potential. Legalization of drug importation in the United States would, therefore, increase manufacturers' reluctance to offer low prices elsewhere. Low-income countries should ban parallel exports as far as is legally possible.

Wealthier countries should also forgo formal and informal referencing to foreign prices, including best-price requirements, because these encourage manufacturers to charge more than they otherwise would in low-income countries.

Manufacturers could be encouraged to grant discounts to low-income countries in the form of confidential rebates paid directly to the ultimate purchaser, while wholesalers and other third-party distributors are supplied at a common price. Achieving differential pricing through confidential discounts prevents other purchasers from demanding matched prices or importing the discounted products. Confidential discounts targeted to programmes for the poor could also be used to reduce prices for low-income populations in countries where market prices are high as a result of wealthy subgroups. Providing differential discounts to health plans is standard practice in the United States, where health plans stimulate competition by demanding discounts in return for an increased market share of a drug through preferred formulary placement[13]. Likewise, purchasers for low-income countries could negotiate volume-related discounts payable by electronic transfer, which would effectively link prices to the country's price sensitivity and pre-empt parallel trade and referencing by other purchasers. Confidential discounting encourages competition, whereas publishing bid prices can lead to price inflexibility and tacit collusion between suppliers.

Confidential discounting to implement differential pricing may conflict with policy pressure for price transparency. However, as purchasing on behalf of developing countries is increasingly done either by governments or large non-governmental organizations (NGOs), such as the United Nations Children's Fund (UNICEF), the Global Fund To Fight AIDS, Tuberculosis and Malaria, and the William J. Clinton Foundation, monitoring could be done by audit by an approved third party. Consistent with the thesis that confidentiality ensures the lowest prices, UNICEF does not publish the supply prices of individual vaccine manufacturers. Other NGOs (for example, Médecins Sans Frontières and the Clinton Foundation) have publicized at least some prices, probably to encourage bigger discounts from other suppliers. But, in general, firms that have significant sales in high-income markets are more likely to grant lower prices to developing countries through confidential discounts, rather than to publicly announce price cuts that could trigger matching demands in other middle- and higher-income countries.

Differential pricing through confidential, negotiated rebates is also flexible and could extend over a broad range of drugs and countries. Broadening access to low-priced drugs is crucial in developing countries, given the large and growing burden of chronic disease for which effective medicines exist but are unaffordable without differential pricing.

The World Trade Organization, through the international TRIPS agreement on intellectual-property rights, permits governments to issue a compulsory licence that requires the patent holder to grant a production licence, usually to a local generic company, in cases of national health emergency. Countries with insufficient manufacturing capacity can also issue a compulsory licence to import any medicine[14] — use is not restricted to national health emergencies. A supplementary statement[15] notes the "shared understanding" that the system would "not be an instrument to pursue industrial or commercial policy objectives".

Compulsory licensing reduces prices only if the licensees have lower costs than the originator firms and if they pass these savings on to consumers. However, labour is a small fraction of production cost, and many multinational R&D-based companies have plants in low-wage countries. If originator firms have higher costs, this more probably reflects costs of compliance with environmental and regulatory requirements, including the US Food and Drug Administration. If originator firms charge higher prices than compulsory licensees mainly to avoid price spillovers (which are not an issue for generic manufacturers), this is better addressed by the measures described earlier to assure market separation, rather than by permitting compulsory licensing.

Any short-term benefit to consumers from compulsory licensing must be weighed against the equally real but less visible negative effects of compulsory licensing on R&D incentives. There is a real risk that compulsory licensing could become more widespread, including by middle- and higher-income countries that seek cheaper drugs and/or increased revenue for local firms, at the risk of undermining incentives to develop new drugs in the longer term. This threat would be reduced by effective differential pricing to keep prices low in low-income countries and moderate in middle-income countries.

> "Broadening access to low-priced drugs is crucial."

'Push' and 'pull' subsidies for R&D

Differential pricing can reconcile R&D incentives with affordability in low-income countries only for drugs with significant sales in high-income countries. For diseases that occur predominantly in low-income countries, revenue from drug sales is not sufficient to attract R&D, hence donor subsidies are necessary. 'Push' subsidies fund R&D directly, usually through specialized public–private partnerships aiming to develop new compounds. Advance market commitments (AMCs) are a type of 'pull' subsidy designed to stimulate R&D: donors make a legally binding commitment to pay a specified price for up to a specified number of units of the drug(s) or vaccine(s), which must meet specified criteria, provided that developing countries commit to use the product and pay their share of the price for a number of years. The G8 leading industrialized nations are developing AMCs for vaccines — a pneumococcal vaccine is a candidate in late-stage development, and a malaria vaccine is a possible early-stage candidate. The appeal of AMCs to donors is that they pay only if firms successfully develop the appropriate new medicines, whereas with push subsidies donors pay in advance and bear the full risk of R&D failure.

It is too soon to tell whether AMCs will be effective in stimulating R&D for neglected diseases and at what cost. An analogy is sometimes drawn between AMCs and the orphan-drug laws, which have been very successful. The 1983 US Orphan Drug Act and the similar 2001 European legislation target diseases that affect fewer than 200,000 patients; they combine push subsidies (through R&D tax credits) with a pull subsidy through a seven-year period of market exclusivity. However, the orphan-drug pull component differs from an AMC in that the orphan-drug supplier is free to set the price, and the seven years' exclusivity bars entry by competitors unless they have a differentiated and superior product. By contrast, firms that compete for an AMC face a price fixed by the donors and a significant volume risk, as a result of both market uncertainty (which developing countries will commit to purchase) and market-share uncertainty (which competitors will enter and at what price).

Realistically, AMCs could accelerate the development and diffusion of vaccines that are already in development, but it will be many years before this mechanism is effective at stimulating investment in early-stage R&D. In the meantime, obtaining orphan status in the United States and European Union may provide some additional revenue even for drugs and vaccines that target diseases of developing countries, through differential pricing to the hospital and traveller markets in high-income countries.

Differential pricing could go a long way towards making drugs that are developed for high-income countries affordable in developing countries, while preserving incentives for R&D. But achieving separate markets and eliminating price spillovers across countries is key to achieving differential pricing in practice. This means that middle- and higher-income countries must forgo parallel trade, referencing their prices to prices in lower-income countries, and demanding best-price equalization — practices that would decline if price discounts to low-income countries were kept confidential. If manufacturers sell to distributors at a uniform price but differentiate final prices to purchasers through confidential rebates, opportunities for parallel trade and external referencing are eliminated. If market separation can be guaranteed so that originator firms can sustain differential pricing, then they could charge prices comparable with those of local generic firms in low-income countries, eliminating the case for compulsory licensing.

As differential pricing alone will not stimulate R&D for medicines to treat diseases that occur only in developing countries, supply-side or demand-side subsidies are necessary for such diseases. The optimal strategy would include the use of both, with AMCs as a demand-side subsidy to stimulate commercialization of promising vaccines and drugs that have demonstrated proof of concept. ■

Patricia M. Danzon is Celia Moh Professor in the Health Care Management Department at The Wharton School, University of Pennsylvania, Philadelphia, USA.

1. Danzon, P. & Towse, A. *Int. J. Health Care Finance Econ.* **3,** 183–205 (2003).
2. DiMasi, J. A., Hansen, R. W. & Grabowski, H. G. *J. Health Econ.* **22,** 151–185 (2003).
3. Ramsey, F. P. *Econ. J.* **37,** 47–61 (1927).
4. Varian, H. R. *Am. Econ. Rev.* **75,** 870–875 (1985).
5. Dumoulin, J. *Int. J. Biotechnol.* **3,** 338–349 (2001).
6. Hausman, J. A. & MacKie-Mason, J. K. *RAND J. Econ.* **19,** 253–265 (1988).
7. Brittan, L. *Making a Reality of the Single Market* SPEECH/92/113 (1 December 1992).
8. Maskus, K. E. *Parallel Imports in Pharmaceuticals: Implications for Competition and Prices in Developing Countries* (World Intellectual Property Organization, Geneva, 2001).
9. Danzon, P. M. & Furukawa, M.F. *Health Affairs* http://content.healthaffairs.org/cgi/content/full/hlthaff.w3.521v1/DC1 (2003).
10. US Congressional Budget Office. *H.R. 2427: The Pharmaceutical Market Access Act of 2003* (2003).
11. Danzon, P. M, Wang Y. R. & Wang, L. *Health Econ.* **14,** 269–292 (2005).
12. Commission on Macroeconomics and Health. *Macroeconomics And Health. Investing in Health for Economic Development* (World Health Organization, Geneva, 2001).
13. Danzon, P. M. *Int. J. Econ. Bus.* **4,** 301–321 (1997).
14. World Trade Organization. *Implementation of Paragraph 6 of the Doha Declaration on theTRIPS Agreement and Public Health* Document WT/L/540 (World Trade Organization, Geneva, 2003).
15. World Trade Organization. *Excerpt from the Minutes of the General Council Meeting 30 August 2003 (paragraph no. 29)* Document WT/GC/M/82 (World Trade Org anization, Geneva, 2003).

The road to recovery

Brazil urgently needs to improve infrastructure for generating pharmaceuticals to alleviate the plight of its poor and marginalized populations, say **Carlos M. Morel** *et al.*

Despite Brazil's strength in basic scientific and medical research, a large proportion of the population still suffers from ill health. Diseases such as tuberculosis and leprosy are highly prevalent in poor populations, and about 46,000 people die each year from infectious diseases. What is going wrong?

As a 'middle-income' country, Brazil needs to cope with diseases and health problems that are prevalent worldwide, such as diabetes, hypertension and obesity. In addition, it has to deal with neglected diseases, such as malaria and leprosy, and some of the world's 'most-neglected' diseases, such as dengue fever, leishmaniasis and Chagas' disease. But few of the scientific discoveries made in Brazil lead to new drugs for these infectious diseases. This is driving the government into the red, with imported medical products costing billions of dollars each year. Moreover, any successes in health-related biotechnology[1–3] are undermined by the poor delivery of healthcare services across the country[4] and by the consistently ineffective implementation of education and industrial policies.

Certainly, there has been no shortage of effort by Brazilian researchers. Since 1990, the number of articles published by researchers from Brazilian institutes has steadily increased (see graphic, below). Similarly, the number of Brazilian patent applications at the US Patent and Trademark Office has also increased during this period. But the ratio of patent applications to research papers is low, suggesting that not enough research is being translated into real products[5].

The path from innovation to application in health systems and services has recently been proposed to involve six components[6,7]. These are carrying out research and development; manufacturing products to appropriate standards; promoting and sustaining domestic markets; promoting and sustaining export markets; creating and implementing systems for intellectual-property management; and designing and implementing systems for the regulation of drugs, vaccines, diagnostics and medical devices. This framework is a useful guide to help analyse the strengths and weaknesses of Brazil's innovation–application process and points to one key issue underlying all these areas in Brazil: a lack of effective infrastructure.

> "There is an urgent need to devise mechanisms, strategies and policies that address key factors affecting the health of the population."

Call for robust policies

Brazil's current policies on science and technology originated in the 1950s with the creation of two government agencies: one for carrying out research, CNPq (initially called the National Research Council, and now known as the National Council for Scientific and Technological Development); and one for training people, CAPES (Coordination for the Training of Human Resources at the University Level). This strategy was inspired by the linear concept of technological development that was prevalent after the Second World War, when the idea of research and development as two distinct activities was born. Basic research was considered to be necessary and sufficient for technological, social and economic improvement. This paradigm had proved useful in developed countries with a strong industrial infrastructure. But in Brazil it culminated in poor communication between academia and industry. As a result, Brazil's industrial policies failed to take into account a growing requirement for more scientific investigation and innovation. This failure also contributed to shaping what has been called a 'passive national learning system' by Eduardo Viotti[8]. This type of system is typical of countries that do not innovate themselves but rely mainly on copying or adapting innovations from elsewhere.

In the absence of long-term, innovation-based industrial policies, successive governments relied on isolated programmes and projects, scattered initiatives, and self-promoting funding agencies to stimulate manufacturing capability in selected areas. Ventures such as the National Immunization Program and the National Self-Sufficiency Program in Immunobiologicals allowed Brazil to become a world leader in terms of immunization strategies and campaigns[9–11], but such success could not be forged in other areas as a result of an overall lack of growth in biotechnology. In this way, Brazil has lagged behind other recently industrialized nations, including South Korea and Singapore.

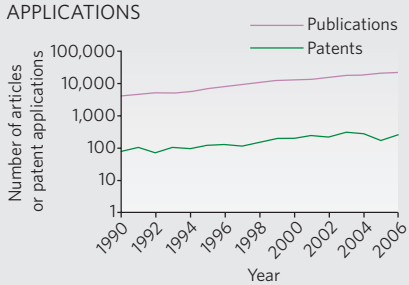
Redressing the imbalance

Moreover, beginning in the 1990s, the Brazilian government adopted policies that introduced further difficulties for local industrial enterprises. For example, Brazil signed up to TRIPS, the international agreement on intellectual-property rules brokered by the World Trade Organization. This agreement permitted a transition period for developing countries, to allow them to adapt to the new legal framework. However, the Brazilian government did not take advantage of this concession, and Brazil's Industrial Property Law quickly became fully compliant with the agreement, in 1996. As a result, several Brazilian biotechnology and pharmaceutical companies — unprepared for the international competition that resulted — did not survive[12]. Today, the outcome of this policy is an annual deficit of US$2 billion in the trade balance for pharmaceutical products (see graphic, opposite) and more than US$3 billion for all medical products[13].

Brazil's substandard educational system and socioeconomic imbalance have also contributed to the generally unfavourable climate for innovation. And this lack of innovation, in turn, has prevented these social problems from being rectified.

But there have been success stories. In the public sector, institutes such as the Oswaldo Cruz Foundation (Fiocruz), in Rio de Janeiro, and the Butantan Institute, in São Paulo, have become well-established producers of immunobiologicals and pharmaceuticals, meeting the requirements of Brazil's Ministry of Health. In addition, the number of graduates and graduate programmes in disciplines relevant to the pharmaceutical industry, particularly chemistry, is increasing[14]. And, in the private sector, companies such as Aché Laboratories, Cristália and Nortec Química are developing and launching new drugs,

BRAZILIAN PUBLICATIONS AND PATENT APPLICATIONS



SOURCES: THOMSON SCIENTIFIC/USPTO, 2007

CCS/FIOCRUZ

Success story: Brazil is the world's largest producer of vaccine against yellow fever.

SOURCES: MDIC, BRAZIL/FEBRAFARMA, 2007

and manufacturing synthetic components of pharmaceuticals.

Given all of these factors, it became clear that the way in which research, development and innovation were being carried out needed to change and that these activities had to be coordinated at the national level, within a robust industrial-policy framework.

Shaping the future

Recently, Brazil started to lay the foundations for better public health. In 2004, the government approved a new legal and regulatory framework, which includes the Industrial, Technological and Foreign Trade Policy (PITCE). This development initiative links diverse support programmes (financial, technological, logistical, commercial and industrial) with the aim of encouraging innovation. Following on from this, in 2005, Brazil introduced the Law of Innovation, which is organized around academic, technological and commercial axes. This law is designed to foster strategic partnerships among universities, technological institutes and businesses, creating a three-pronged approach to innovation.

In addition, a Biotechnology Development Policy was developed this year, providing a broad, structured industrial policy with long-term goals involving the public and private sectors. The policy wisely spreads responsibility across several branches of the government, unlike the earlier isolated and disparate initiatives.

The recent policy and legal changes should help to generate new drugs to treat neglected diseases, by bridging the gap between basic research and drug development. For example, the Department of Science and Technology, which was created as part of Brazil's Ministry of Health in 2000, joined forces in 2006 with the Ministry of Science and Technology (through CNPq) to tackle six neglected and most-neglected diseases: dengue fever, Chagas' disease, leprosy, malaria, tuberculosis and the various forms of leishmaniasis. Together, they have invested US$10 million in 76 peer-reviewed projects, as part of a pilot research-and-development programme in 2007–08. The programme builds on existing international networks in which scientists actively collaborate in the study of neglected diseases (see 'Collaborative research networks', overleaf). And it aims to strengthen capacity for research on neglected diseases, particularly in regions of Brazil where these diseases are endemic.

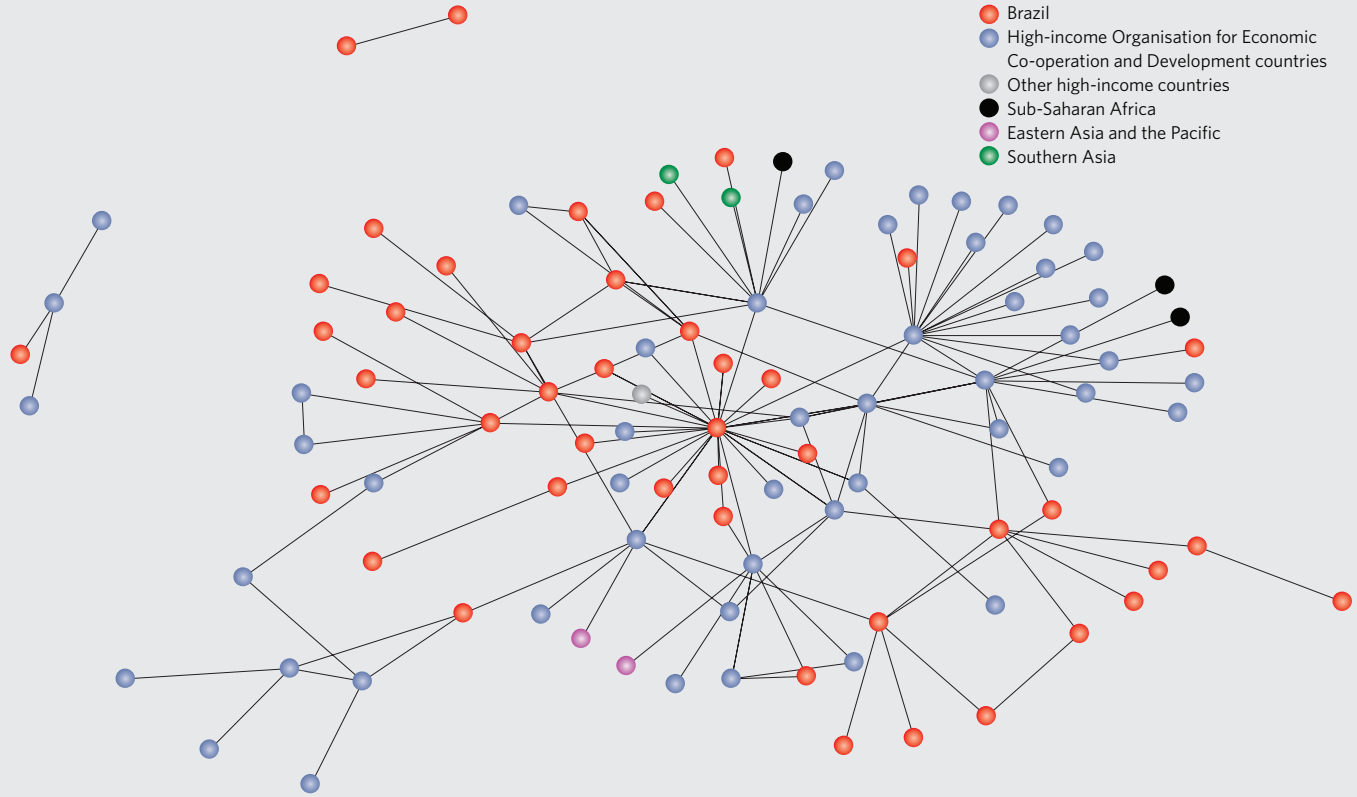
A move towards manufacturing

Moving from research on neglected diseases to the development and manufacture of products that prevent, treat and control them will not be an easy task for Brazil. Areas that traditionally have required little technology (such as vaccine development), giving the public sector an advantage, are being overtaken by technologically intensive and expensive production processes. So there is an urgent need to devise new mechanisms, strategies and policies, within the framework of PITCE, that address key factors affecting the general health of the population[6,7].

These strategies should include the deployment of public-sector funds to support public and private manufacturers of treatments for neglected diseases, as well as agencies that manage intellectual-property rights and oversee health regulations.

There are several private initiatives in Brazil

COLLABORATIVE RESEARCH NETWORKS



In addition to formal initiatives to promote collaboration, there are already less formal networks between Brazilian researchers and their colleagues worldwide. These networks help to accelerate the progress of research and development, through the sharing of ideas and resources[15].

Here, we focus on research into a single neglected disease to illustrate the diversity of existing networks. The graphic represents the network of institutes that connects Brazilian scientists studying leprosy with other scientists around the world. The network was generated by extracting author affiliations from articles about leprosy that were published by Brazilians between 2001 and 2005, according to Thomson Scientific. And the position of each institute in the network was determined using the social-networking software UCINET[16]. Each circle represents a particular institute, and each line represents at least one collaboration between these institutes, defined as co-authorship on a scientific article. Institutes are colour coded according to geographical and economic categories designated by the World Bank. It should be noted that red indicates Latin America and the Caribbean, but in this example, all of the institutes in this category are Brazilian.

It is clear from this example that Brazilian researchers are an integral part of a broad international research network that involves both developed and developing countries. It is also evident that Brazilian institutes can have central roles in collaborative research networks that investigate neglected diseases. **C.M.M.**

that stimulate collaboration between academia and industry on neglected diseases. In addition, since the late 1990s, several global public–private partnerships have been set up to develop drugs, vaccines and diagnostics for neglected diseases (see pages 164 and 166). These include the Malaria Vaccine Initiative, the Drugs for Neglected Diseases Initiative and the Foundation for Innovative New Diagnostics. Brazil should be a welcome contributor to such partnerships, whose rewards could markedly enhance the pharmaceutical production on which the health of Brazil's people depends.

A welcome move in this direction was made this year. Fiocruz and its Centre for Technological Development in Health signed a collaborative agreement with the biotechnology company Genzyme, through its Humanitarian Assistance for Neglected Disease project, to work on the research and development of drugs for 17 neglected diseases. Such strengthening of international collaborations and public–private partnerships, as well as improvements in industrial and educational policies, is crucial for Brazil to forge an effective national innovation system and continue on the road to recovery. ■

Carlos M. Morel is director of the Centre for Technological Development in Health, Fiocruz, Rio de Janeiro, Brazil. José R. Carvalheiro is vice-president for research and development at Fiocruz. Carmen N. P. Romero is a science and technology analyst on the Innovation Project at Fiocruz. Eduardo A. Costa is director of the Medicines and Drugs Technology Institute, Fiocruz. Paulo M. Buss is the president of Fiocruz.

1. Levi, G. C. & Vitoria, M. A. *AIDS* **16,** 2373–2383 (2002).
2. Ferrer, M., Thorsteinsdottir, H., Quach, U., Singer, P. A. & Daar, A. S. *Nature Biotechnol.* **22,** DC8–DC12 (2004).
3. Olive, J. M., Risi, J. B. & de Quadros, C. A. *J. Infect. Dis.* **175,** S189–S193 (1997).
4. Almeida, C. [in Spanish with English abstract] *Cad. Saúde Pública* **18,** 905–925 (2002).
5. Bernardes, A. T. & da Motta e Albuquerque, E. *Res. Pol.* **32,** 865–885 (2003).
6. Morel, C. *et al. Innov. Strategy Today* **1,** 1–15 (2005).
7. Mahoney, R. T., Krattiger, A., Clemens, J. & Curtiss, R. *Vaccine* **25,** 4003–4011 (2007).
8. Viotti, E. B. *Technol. Forecast. Soc.* **69,** 653–680 (2002).
9. Temporão, J. G. [in Portuguese with English abstract] *Hist. Cienc. Saúde Manguinhos* **10,** 601–617 (2003).
10. Gadelha, C. & Azevedo, N. [in Portuguese with English abstract] *Hist. Cienc. Saúde Manguinhos* **10,** 697–724 (2003).
11. Buss, P. M., Temporão, J. G. & Carvalheiro, J. R. *Vacinas, Soros e Imunizações no Brasil* [in Portuguese] (Editora Fiocruz, Rio de Janeiro, 2005).
12. Basu, P. *Nature Biotechnol.* **23,** 13–15 (2005).
13. Gadelha, C. A. G. [in Portuguese with English abstract] *Rev. Saúde Pública* **40,** 11–23 (2006).
14. Gama, A. A. S., Cadore, S. & Ferreira, V. F. [in Portuguese with English abstract] *Quím. Nova* **26,** 618–624 (2003).
15. Morel, C. M. *et al. Science* **309,** 401–404 (2005).
16. Borgatti, S. P., Everett, M. G. & Freeman, L. C. UCINET 6.0 Version 1.00 (Analytic Technologies, Natick, 1999).

ARTICLES

MicroRNA control of Nodal signalling

Graziano Martello[1], Luca Zacchigna[1], Masafumi Inui[1], Marco Montagner[1], Maddalena Adorno[1], Anant Mamidi[1], Leonardo Morsut[1], Sandra Soligo[1], Uyen Tran[2], Sirio Dupont[1], Michelangelo Cordenonsi[1], Oliver Wessely[2] & Stefano Piccolo[1]

MicroRNAs are crucial modulators of gene expression, yet their involvement as effectors of growth factor signalling is largely unknown. Ligands of the transforming growth factor-β superfamily are essential for development and adult tissue homeostasis. In early *Xenopus* embryos, signalling by the transforming growth factor-β ligand Nodal is crucial for the dorsal induction of the Spemann's organizer. Here we report that *Xenopus laevis* microRNAs miR-15 and miR-16 restrict the size of the organizer by targeting the Nodal type II receptor Acvr2a. Endogenous miR-15 and miR-16 are ventrally enriched as they are negatively regulated by the dorsal Wnt/β-catenin pathway. These findings exemplify the relevance of microRNAs as regulators of early embryonic patterning acting at the crossroads of fundamental signalling cascades.

MicroRNAs (miRNAs) represent a new class of non-coding regulatory molecules, the diverse functions of which are just beginning to be explored[1,2]. Although computational predictions suggest that miRNAs can regulate a substantial fraction of the genome[3,4], little is known about their role as transducers or downstream effectors of growth factor signalling.

Cytokines of the transforming growth factor-β (TGF-β) family are prominent signals regulating cell behaviours in a variety of cellular contexts[5]. TGF-β ligands promote oligomerization of cognate serine/threonine type I and type II receptors leading, intracellularly, to the activation of the Smad transducers[5]. During embryonic development, this signalling cascade is activated by Nodal to mediate the induction and patterning of the mesoderm germ layer[6]. Nodal activity, however, is not even along the dorsoventral axis of the developing embryo: the Spemann's organizer is induced on the dorsal side by an asymmetric wave of Nodal signalling that can be visualized by a peak in the level of phosphorylated (p)-Smad2 at the late blastula stage[6–8].

The establishment of embryonic dorsoventral polarity can be ultimately traced back to fertilization: sperm entry breaks the symmetry of the egg, causing the accumulation of Wnt/β-catenin signalling on the future dorsal side[9,10]. The signal initiated by β-catenin is then translated, by unknown mechanisms, into the Nodal gradient, such that embryos with a maternal Wnt/β-catenin deficiency lack the dorsal accumulation of p-Smad2 (ref. 7) and do not develop the organizer tissue[11,12]. Some Nodal-related ligands are expressed with a dorsal-to-ventral gradient during gastrulation[13]; however, the most potent mechanism for maintaining and upregulating Nodal expression is Nodal signalling itself, through a Smad2/FAST-1-dependent autoregulatory loop that feeds on Nodal transcription[13–16]. In other words, higher Nodal expression on the dorsal side may be considered an amplification of an early bias in Nodal signalling[6], but does not inform us of the mechanisms that initiate this asymmetry. Here we searched for miRNAs affecting Nodal signalling and identified miR-15 and miR-16 as molecules that translate the early asymmetry in Wnt/β-catenin into the generation of a gradient of Nodal responsiveness by partitioning the Nodal receptor Acvr2a along the dorsoventral axis.

miR-15 and miR-16 limit Nodal and activin responses

To test whether miRNAs control Nodal/TGF-β signalling during mesoderm development, we monitored how raising the level of the endogenous pool of miRNAs expressed in early *Xenopus* embryos impinges on Nodal activity. To this end, we purified total miRNA from embryos at the gastrula stage and microinjected 3 ng of this preparation together with Nodal (*Xnr-1*, also called *nr1-A*) messenger RNA into 2-cell-stage *Xenopus* embryos. Nodal gene responses were monitored by the co-injected *Mix.2*-lux reporter, a sensitive and well-established read-out of Nodal/TGF-β-dependent Smad activation[17,18]. As shown in Fig. 1a, embryos with raised miRNA levels exhibited a marked inhibition of Nodal activity. This effect seemed to be specific because, as a control, we injected up to 12 ng of a custom-designed library of 172 small-interfering (si)RNAs against a collection of human genes and found no effect on *Mix.2*-lux induction (lane 3).

MicroRNAs negatively regulate gene expression by annealing to the 3′ untranslated region (UTR) of a target mRNA via their 'seed' sequence (5′ nucleotides 2–8)[1]. We then used Pictar[19] and miRanda[20] computational tools to search for miRNA-binding sites in the 3′ UTR of the core components of the Nodal signalling pathway (ligands, receptors, co-receptors, Smads and nuclear cofactors[5]). From this analysis, the miR-15 family and 3′ UTR of *Acvr2a* (also called *ActRIIA*), a type II receptor for Nodal and activin ligands[21] (Supplementary Fig. 1), attracted our attention for three reasons: (i) the binding site is evolutionarily conserved from amphibians to humans; (ii) mature miR-16 is an abundant miRNA expressed in early *Xenopus* embryos[22]; and (iii) there are two predicted miR-15 and miR-16 binding sites in *Acvr2a* mRNA, suggesting a cooperative binding and thus a biologically effective interaction[1].

Several experiments were carried out to validate the hypothesis that the miR-15 family targets *Acvr2a*. First, we generated two reporter constructs containing either the wild-type or a mutated miR-15-binding site of *Acvr2a* 3′ UTR cloned downstream of the luciferase open reading frame (ORF)[23] (Acvr2a-WT and Acvr2a-mut, respectively) (Supplementary Fig. 2a). We then compared the activity of these two reporters in *Xenopus* embryos and human HepG2 cells, and found that Acvr2a-WT displayed significantly reduced expression compared to Acvr2a-mut, as expected if the endogenous miR-15 and miR-16 were pairing to the predicted target site (Fig. 1b, compare lanes 1 and 4; see also Supplementary Fig. 2b). Overexpression of the mature miR-15 or its primary precursor (pri-miR-15/16) further inhibited Acvr2a-WT expression but had no

[1]Department of Histology, Microbiology and Medical Biotechnologies, Section of Histology and Embryology, University of Padua, viale Colombo 3, 35126 Padua, Italy. [2]Departments of Cell Biology & Anatomy and Genetics, LSU Health Sciences Center, 1901 Perdido Street, New Orleans, Louisiana 70112, USA.

effect on the mutant reporter (Fig. 1b, lanes 2 and 3 versus lanes 5 and 7; see also Supplementary Fig. 2b). In contrast, a mutant miR-15 (containing three mutations in the seed sequence), or an unrelated pri-miRNA, had no effect (Fig. 1b, lanes 6 and 8). Notably, miR-15 downregulated endogenous Acvr2a protein expression, without affecting *Acvr2a* mRNA levels (Supplementary Fig. 2d, e).

By targeting *Acvr2a*, miR-15 and miR-16 should inhibit Nodal responsiveness. To test this, pri-miR-15/16 or mature miR-15 was injected together with the *Mix.2*-lux reporter into the marginal zone of 4-cell-stage *Xenopus* embryos; that is, into the tissue displaying high levels of Nodal-dependent Smad activation[6]. As shown in Fig. 1c, miR-15 and miR-16 inhibits endogenous Nodal signalling. Moreover, the response to activin was inhibited in human HepG2 cells transfected with wild-type miR-15 (Fig. 1d). Notably, raising the level of Acvr2a (upon transfection of an expression construct rendered miRNA-insensitive by deleting the 3′ UTR) rescues the inhibitory effects of miR-15 on activin signalling (Fig. 1d, compare lanes 2 and 3); moreover, transfection of siRNA targeting *Acvr2a* in HepG2 cells mimics miR-15 overexpression (Fig. 1d, lanes 2 and 5). We conclude from these results that miR-15 and miR-16 attenuate Nodal/activin responsiveness by targeting *Acvr2a* (see also Supplementary Fig. 2f–h).

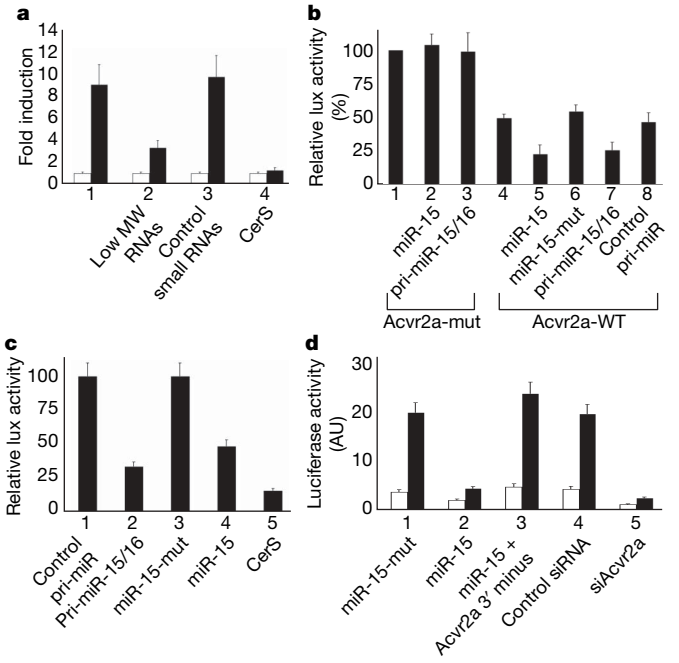


Figure 1 | miR-15 and miR-16 control Nodal/activin responsiveness by acting as inhibitors of Acvr2a expression. **a**, *Xenopus* embryos were injected in the animal pole with the *Mix.2*-lux reporter with (black bars) or without (white bars) 30 pg of *Xnr-1* mRNA. Nodal activity (lane 1) is antagonized by co-injection with 3 ng (lane 2) of endogenous miRNAs purified from stage 10 embryos. Low molecular weight (low MW) RNAs were <100 bp. Co-injection of a complex mix of small RNAs (12 ng) has no effect (lane 3). Injection of the Nodal antagonist CerS[25] serves as a positive control. **b**, The predicted miR-15 and miR-16 binding site in the 3′ UTR of *Acvr2a* mRNA is responsive to endogenous miRNAs as well as to overexpressed mature miR-15 (15 ng) or pri-miR-15/16 (7 ng). Embryos were radially injected in the marginal zone with 40 pg of reporter plamids (Supplementary Fig. 2a) alone (lanes 1 and 4) or in combination with the indicated miRNAs. miR-15 and miR-16 mature levels are shown in Supplementary Fig. 2c. **c**, The transcription of the *Mix.2*-lux reporter injected in the marginal zone of 4-cell-stage embryos relies on endogenous Nodal (compare lanes 1 and 3 with lane 5, showing the basal *Mix.2* transcription in embryos injected with CerS). Mature miR-15 as well as pri-miR-15/16, but not the unrelated pri-miR-127 (control) or a mutant miR-15 (miR-15-mut), inhibit this response. **d**, Transfection of wild-type mature miR-15 inhibits the responsiveness to activin (1 ng ml^{-1}) in HepG2 cells. This effect is specific, as miR-15 overexpression is inactive against signalling initiated by TGF-β1 or BMP (Supplementary Fig. 2f–h). Co-transfection of *Acvr2a* 3′ UTR-free construct rescues miR-15 inhibitory activity. Transfection of siRNA for *Acvr2a* phenocopies miR-15 overexpression. Untreated, white bars; activin, black bars. AU, arbitrary units. Graphs show representative experiments each carried out at least three times (**a**–**c**) or twice (**d**) independently, with comparable results. Error bars show s.d.

miR-15 and miR-16 set the size of the organizer

In vertebrates, Nodal signalling is crucial for the induction and patterning of the mesoderm germ layer[6]. Before gastrulation, an asymmetric peak of Nodal signalling on the dorsal side of the embryo induces the Spemann's organizer[7,8,13]. To unveil the biological function of miR-15 and miR-16 in these events, we first carried out gain-of-function studies in *Xenopus* by analysing the expression of Nodal targets in miR-15-injected embryos. Misexpression of miR-15 downregulated the organizer markers *chordin*, *pintallavis* and *Xantivin*, and negatively regulated mesoderm induction, reducing the strength and width of *Xbra* (Fig. 2 and Supplementary Fig. 3a). However, the *Xbra* staining was not erased in miR-15-injected embryos, and mesodermal derivatives, such as somites, are formed in these embryos (data not shown), in agreement with the phenotype of mice with incomplete genetic inactivation of Acvr2 receptors[21]. Confirming that the defects induced by miR-15 overexpression are caused by defective Acvr2a signalling, normal *chordin* and *pintallavis* expression was effectively rescued by co-injecting miR-15 with an *Acvr2a* mRNA lacking its 3′ UTR (3′ minus); expression was less efficiently rescued by co-injection with wild-type *Acvr2a* mRNA (3′ plus) (Fig. 2a and Supplementary Fig. 3b). We conclude from these experiments that miR-15 operates by antagonizing Acvr2 signalling in early

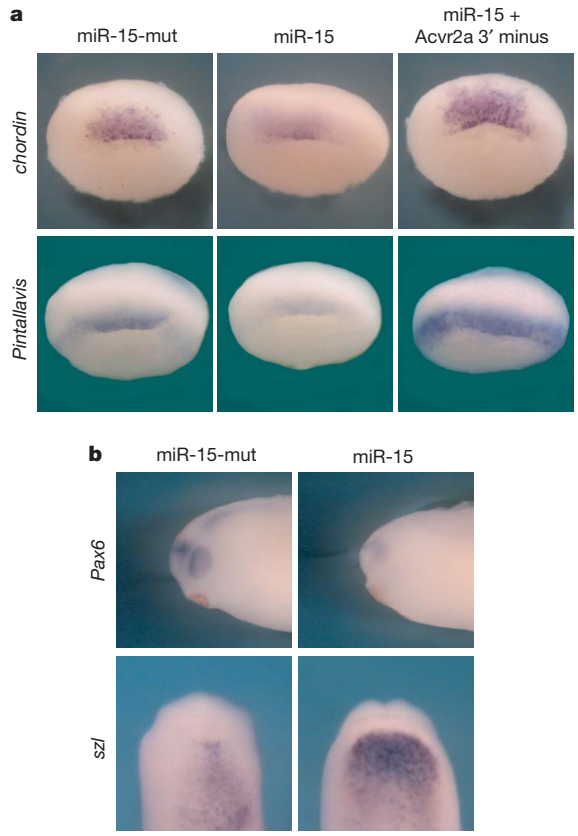


Figure 2 | miR-15 antagonizes Spemann's organizer development. **a**, Whole-mount *in situ* hybridizations (dorsal side is up) for early organizer markers in embryos overexpressing miR-15 (*chordin* and *Pintallavis*); mutant miR-15 was injected as a control. Co-injection of *Acvr2a* 3′ UTR-free mRNA (150 pg) rescues organizer gene expression. **b**, Later phenotypic effects of miR-15 misexpression. Note anterior deficiencies (*Pax6*) and ventralization (*szl*). See Supplementary Table 1 for statistics.

embryos. Accordingly, Nodal signalling initiated downstream of the receptor after injection of *Smad2* mRNA is insensitive to miR-15 overexpression (Supplementary Fig. 3c).

We next extended the analysis to embryos at later stages. *In situ* hybridizations for *sonic hedgehog* and *chordin* show that raising miR-15 levels leads to defective formation of the most anterior segment of the axial mesoderm, namely, the earliest organizer derivative (Supplementary Fig. 3d); in contrast, formation of the notochord was marginally affected. These data are consistent with the requirement of high levels of Nodal signalling in prechordal mesoderm, whereas lower levels are sufficient for notochord differentiation[6]. The prechordal plate has key roles in patterning the anterior neural plate: accordingly, miR-15-injected embryos show reduced expression of the anterior neural markers *Pax6* and *XBF-1* (also called *foxg1b-a*), and, ultimately, have smaller heads (Fig. 2b, Supplementary Fig. 3e, and data not shown). Further read-out of defective organizer formation was the expanded expression of the ventral marker *sizzled* (*szl*) (Fig. 2b).

To ascertain the relevance of miR-15 and miR-16 for Nodal signalling *in vivo*, we carried out loss-of-function studies. To this end, we used 2′-*O*-methyl antisense inhibitory oligonucleotides (anti-miRNAs), which have been shown to inhibit miRNA activity[23]. We first designed two anti-miRNAs, one against miR-15 and one against miR-16 (hereafter called anti-miR-15/16), and tested them for their ability to inhibit endogenous miR-15 and miR-16 function. We found that, compared to control anti-miRNA, microinjection of anti-miR-15/16 alleviated the repression on the 3′ UTR of *Acvr2a* by endogenous miR-15 and miR-16 (Supplementary Fig. 4a). Crucially, loss of miR-15 and miR-16 function raised the steady-state levels of Acvr2a protein (Supplementary Fig. 4b). Next, we tested whether inactivation of miR-15 and miR-16 impinged on Nodal signalling. Embryos with inactivated miR-15 and miR-16 function showed an enhanced activation of the *Mix.2*-lux promoter, indicating that endogenous miR-15 and miR-16 restrict Nodal responsiveness in early embryos (Fig. 3a). Further supporting this notion, we found that knockdown of miR-15 and miR-16 or overexpression of *Acvr2a* mRNA (3′ UTR minus) rescues the inhibition of Nodal signalling caused by injected total embryonic miRNAs (Supplementary Fig. 4c).

Because ectopic expression of miR-15 inhibited dorsal development, we investigated whether the inactivation of endogenous miR-15 and miR-16 favours the formation of Spemann's organizer. For this purpose, 4-cell-stage embryos were radially injected with anti-miR-15/16 in the marginal zone and analysed at the gastrula stage. Inactivation of miR-15 and miR-16 caused a remarkable expansion of organizer tissue, as revealed by *in situ* hybridization with several markers; as a control, we used embryos injected with an unrelated anti-miRNA or with an anti-miRNA targeting the 'passenger' strand of miR-16 (Fig. 3b and Supplementary Fig. 5).

We then sought for independent evidence of the relevance of miR-15 and miR-16 in embryonic patterning. As an alternative to anti-miRNA, we designed four morpholino oligonucleotides targeting the miR-15 family members, as these reagents have been recently shown to block miRNA function[24]. Injection of a mix of miR-15 and miR-16 morpholino oligonucleotides (35 ng each) led to an expansion of Spemann's organizer gene expression, in agreement with the results obtained with 2′-*O*-methyl oligonucleotides (Supplementary Fig. 6). These effects are due to an increased responsiveness to Nodal, as they are reversed, dose-dependently, by the expression of the Nodal antagonist cerberus-short (CerS)[25] and mimicked by Acvr2a overexpression (Supplementary Fig. 7). Accordingly, the activity of *Mix.2*-lux is enhanced by miR-15 and miR-16 morpholino oligonucleotides (Supplementary Fig. 6). Taken together, the data indicate that the size of the Spemann's organizer is limited by miRNAs of the miR-15 family.

miR-15 and miR-16 are ventrally enriched

Because miR-15 is required to limit Nodal responses, we reasoned that uneven localization of miR-15 and miR-16 might be involved in patterning Nodal activity in *Xenopus* embryos (Fig. 4a). We therefore measured the levels of mature miR-15 and miR-16 in the ventral and dorsal marginal zone both directly, by polymerase chain reaction with reverse transcription (RT–PCR)[26], and indirectly, by comparing the activity of the miR-15 reporters (Acvr2a-WT and Acvr2a-mut). As shown in Fig. 4b, c, miR-15 and miR-16 levels

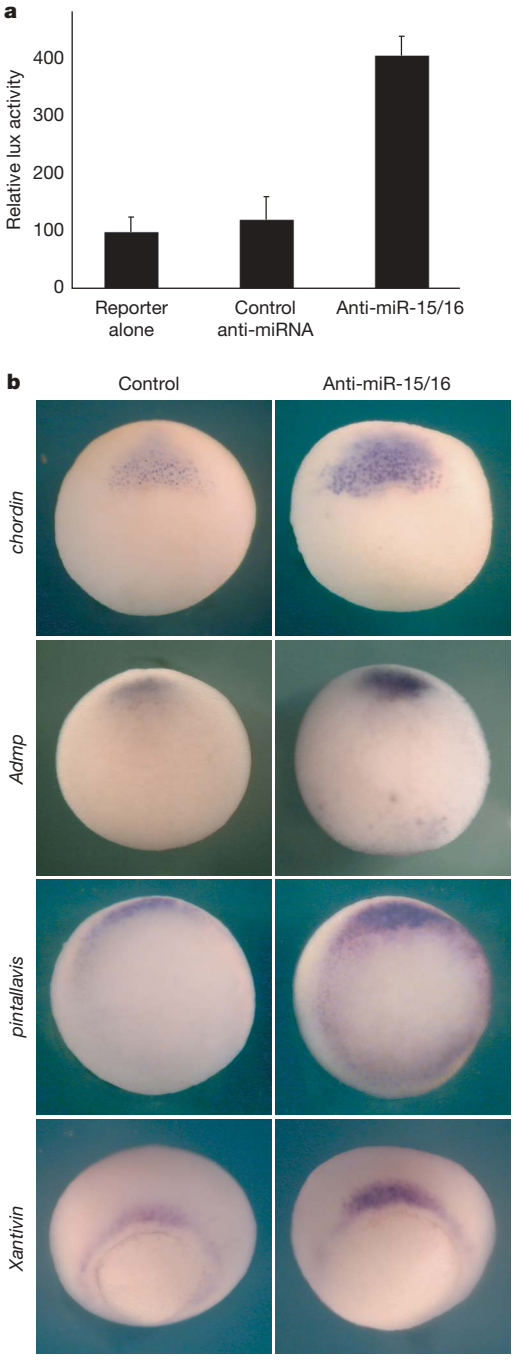


Figure 3 | miR-15 and miR-16 are required to set the size of the organizer by limiting Nodal responsiveness. a, **b**, Effect of miR-15 and miR-16 functional inactivation with anti-miRNA reagents (14 ng; a 1:1 mix of anti-miR-16 and anti-miR-15 injected radially at the 4-cell stage) on endogenous Nodal signalling (**a**), and on embryonic organizer development (**b**). **a**, Loss of miR-15 and miR-16 increases *Mix.2*-lux transcription (as in Fig. 1c). Graph shows representative experiments each carried out at least three times independently, with comparable results. Error bars show s.d. **b**, Injection of anti-miR-15/16 leads to lateral expansion and upregulation of organizer markers (*chordin*, *Admp*, *pintallavis*, *Xantivin*). Statistics of the observed phenotypes are in Supplementary Table 1.

and activity were more abundant on the ventral side of the embryo. Notably, the asymmetry of mature miR-15 and miR-16 reflects a complementary gradient of endogenous Acvr2a protein expression, as revealed by western blotting for Acvr2a in dorsal, lateral and ventral marginal zone explants, and by immunohistochemistry (Fig. 4d and Supplementary Fig. 8). These data suggest that, by inhibiting Acvr2a protein expression on the ventral side, miR-15 and miR-16 might render dorsal cells more responsive to Nodal ligands. To validate this hypothesis, we compared the levels of endogenous p-Smad2 in dorsal and ventral marginal zone explants from wild-type embryos and from embryos injected with pri-miR-15/16 or anti-miR-15/16. As shown in Fig. 4e, dorsal accumulation of p-Smad2, evident in wild-type explants, was lost in miR-15-injected explants (lanes 1 and 2 versus 3 and 4). In contrast, blockade of miR-15 function increases ventral and, to a lesser extent, dorsal p-Smad2 levels (lanes 5 and 6). Together, these results suggest a model in which a ventral bias in miR-15 and miR-16 expression imposes a graded expression pattern on the Nodal receptor, resulting in increased Nodal responsiveness on the dorsal side (Fig. 4f); this mechanism locks the Spemann's organizer gene expression programme in place.

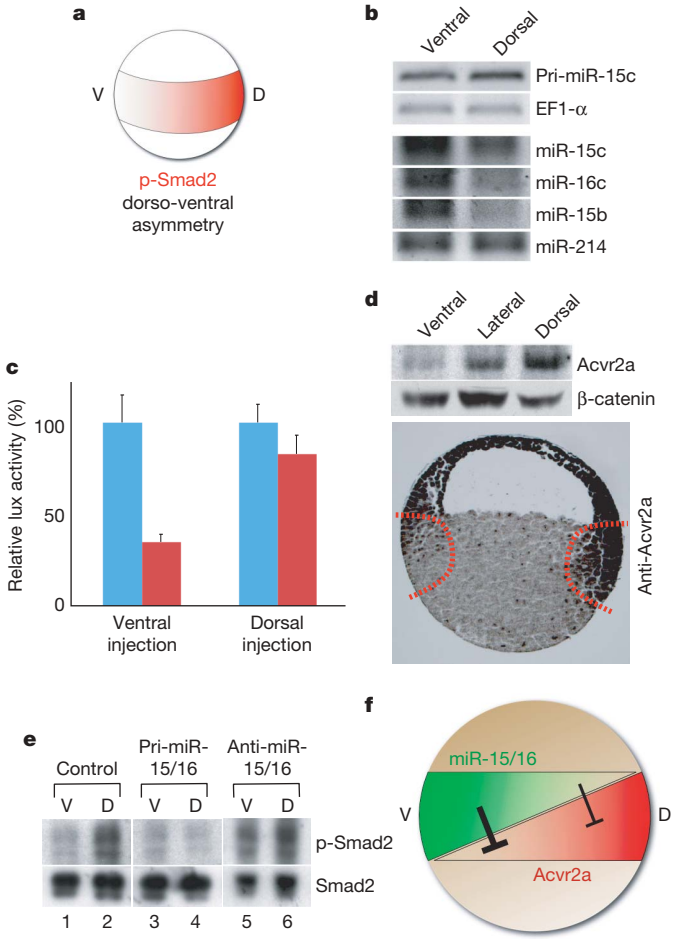


Figure 4 | miR-15 and miR-16 and Acvr2a show complementary domains of expression. **a**, Schematic diagram showing the asymmetric dorsal bias in p-Smad2 distribution in *Xenopus* embryos at the late blastula stage[7,8]. D, dorsal side; V, ventral side. **b**, RT–PCR detecting primary transcript or mature miR-15 family members in dorsal and ventral marginal zone explants from stage 9 embryos. Endogenous mature miRNA levels are ventrally enriched, but pri-miR-15/16c is evenly distributed. miR-214 and EF1-α serve as loading controls. **c**, Wild-type (red bars) and mutant (blue bars) miR-15 reporters from *Acvr2a* 3′ UTR were injected either dorsally or ventrally in regularly cleaving 4-cell-stage embryos. Endogenous miR-15 and miR-16 activity is ventrally enriched. Graph shows representative experiments carried out twice with comparable results; error bars show s.d. **d**, Top: western blotting for endogenous Acvr2a protein using lysates of dorsal, lateral and ventral explants (from stage 10 minus embryos). Detection of β-catenin serves as a loading control. Bottom: immunohistochemical localization of endogenous Acvr2a protein in early *Xenopus* embryos (see also Supplemental Fig. 8). **e**, Western blotting for endogenous p-Smad2 in dorsal (D) and ventral (V) halves[8] (at stage 9.5). Detection of total Smad2 serves as a loading control. **f**, Model showing the complementary asymmetries in miR-15 and miR-16 and Acvr2a distributions.

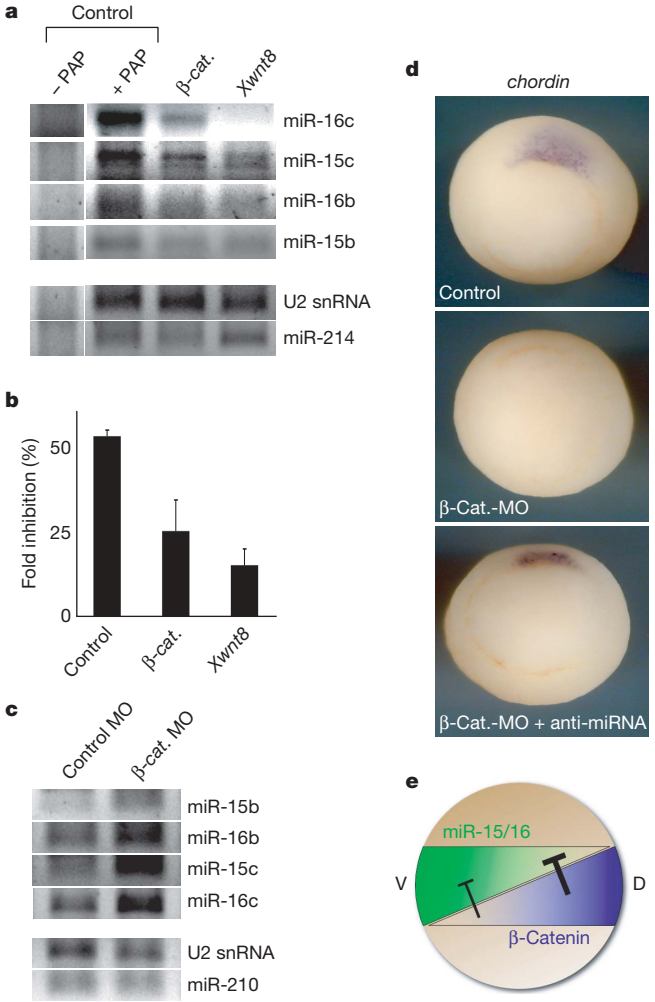


Figure 5 | miR-15 and miR-16 are negatively regulated by Wnt/β-catenin signalling. **a**, RT–PCR detecting mature miR-15 family members in embryos injected with 200 pg of *β-catenin* and 100 pg of *Xwnt-8* mRNAs. Lanes marked control were not injected. miR-15a was not detectable (data not shown). Loading controls: U2 is a small nuclear RNA and miR-214 is an unrelated miRNA. Control samples without PAP (PolyA-polymerase) ensure the amplification of mature miRNAs. **b**, β-Catenin or Wnt misexpression inhibits the activity of endogenous miR-15 and miR-16, as monitored by means of the *Acvr2a* 3′ UTR reporters (see Methods). The graph shows representative experiments carried out twice with comparable results; error bars show s.d. **c**, RT–PCR detecting mature miR-15 family members in embryos with reduced β-catenin levels after injection of β-catenin morpholino oligonucleotides (MO) at the 2-cell stage (50 ng). RT–PCRs for U2 and miR-210 are loading controls. **d**, *In situ* hybridization for the organizer marker *chordin*. Top, control embryo; middle, embryo radially injected with β-catenin morpholino oligonucleotide at the 2-cell stage (72% of affected embryos with no *chordin* staining, $n = 43$); bottom, embryo first radially injected with β-catenin morpholino oligonucleotide as above and then receiving a single dorsal injection of anti-miR-15/16 (8 ng) at the 4-cell stage (71% of rescued embryos, $n = 24$). **e**, Model of the complementary asymmetries in early Wnt/β-catenin signalling and miR-15 and miR-16 localization.

Wnt signal inhibits miR-15 and miR-16 expression

Maternal Wnt/β-catenin signalling is also essential for Spemann's organizer formation, acting upstream of Nodal[6,9]. We therefore considered the possibility that miR-15 and miR-16 might link the Wnt and Nodal pathways. To test this hypothesis, we monitored the effect of gain- and loss-of-function of Wnt/β-catenin signalling on endogenous miR-15 and miR-16. Overexpression of *Xwnt-8* or *β-catenin* mRNA inhibits the expression of mature miR-15 and miR-16 isoforms, and, coherently, reduces the capacity of endogenous miR-15 and miR-16 to inhibit the activity of the Acvr2a reporters (Fig. 5a, b). In contrast, knockdown of β-catenin by radial injection of anti-β-catenin morpholino oligonucleotides increased the level and activity of endogenous miR-15 and miR-16 (Fig. 5c and data not shown). Interestingly, gain or loss of Wnt/β-catenin signalling has a minimal effect on the levels of immature pri-miR-15/16c (Supplementary Fig. 9a), suggesting that the maternal Wnt/β-catenin pathway regulates, primarily, miR-15/16 maturation rather than its transcription. This notion is also supported by the even distribution of pri-miR-15/16c along the dorsal–ventral axis (top of Fig. 4b and Supplementary Fig. 9b), which contrasts with the asymmetric localization of the corresponding mature miRNAs.

To understand further the mechanism underlying the regulation of miR-15 and miR-16 by Wnt signalling, we took advantage of the early blockade of zygotic transcription that starts in *Xenopus* only after mid-blastula. We found that *Xwnt-8* overexpression or loss of β-catenin is able to modulate levels of mature miR-15 and miR-16 before mid-blastula (Supplementary Fig. 9c), and thus, formally, in the absence of transcriptional intermediates. This at least suggests an attractive possibility: that Wnt signalling can control miR-15 and miR-16 maturation directly, perhaps through a protein complex controlled by—or containing—β-catenin.

Finally, we investigated whether β-catenin-mediated repression of miR-15 and miR-16 is instrumental for formation of the Spemann's organizer. To this end, microinjection of anti-miR-15/16 into a single blastomere of β-catenin-depleted embryos was carried out and the embryos were assayed for expression of the organizer marker *chordin* at gastrula stage; this experimental set-up allowed us to visualize whether adding back the inhibition of miR-15 and miR-16 is sufficient to complement loss of β-catenin. Notably, loss of miR-15 and miR-16 function rescued *chordin* expression, suggesting that miR-15 and miR-16 are epistatic to β-catenin (Fig. 5d). Together, these results suggest that one key mechanism by which β-catenin controls the intensity and spatial pattern of Nodal responsiveness is through downregulation of miR-15 and miR-16 expression (Fig. 5e).

Discussion

We report that Nodal/activin signalling is controlled by the miR-15 family, representing the first identification of endogenous miRNAs targeting the TGF-β signalling cascade. In contrast to the known regulators of Nodal signalling, miR-15 and miR-16 are enriched on the ventral side of the *Xenopus* embryo, thereby generating a dorsal bias in Acvr2a expression. This regulation is required for the proper formation of the Spemann's organizer. Gain of miR-15 function restricts the size of the organizer, an effect that is consistent with the phenotype of *Xenopus* embryos with reduced Nodal function[6] and those of mice carrying genetic inactivation of *Acvr2a*[21] or Nodal hypomorphic mutants[6]. In contrast, inhibition of miR-15 and miR-16 leads to a lateral expansion and a reinforcement of organizer markers, an effect mimicked by Acvr2a overexpression; a similar phenotype was previously reported in vertebrate embryos lacking the Nodal antagonists *Lefty1* and *Lefty2* or with stabilized Nodal receptors[15,27–29].

Interestingly, the interaction between miR-15 and Acvr2a is conserved from amphibians to humans; however, this interaction is not conserved in teleosts (data not shown). This may explain the absence of obvious defects in early body plan specification in zebrafish embryos lacking Dicer[30], opposed to the very early requirement of Dicer in mammals[31,32].

It has been proposed that one role of miRNAs is to prevent illegitimate translation of mRNAs in domains where these genes are already transcriptionally repressed, thus maintaining tissue identity[1]. However, this principle may not be applicable to embryonic tissues, which contain pluripotent cell populations and where many cell-fate decisions are still plastic and reversible. Our results offer a new mechanism by which miRNAs can specify cell lineage, that is, by partitioning a growth factor receptor. In addition, we provide biological and biochemical evidence that expression of miR-15 family members is under the negative control of Wnt/β-catenin signalling. This cross-talk helps to address a long-standing issue in developmental biology: how, in embryos, the initial asymmetry in β-catenin that ensues after fertilization is translated into a graded Nodal activity[6,9], essential for proper Spemann's organizer formation. It is tempting to speculate that this is an example of a more general principle and that other miRNAs may serve to break the equivalence of otherwise homogeneous cell populations at multiple points during development. Addressing these issues must await a better functional characterization of the role of the 'miRNome' in signalling pathways.

METHODS SUMMARY

For RT–PCR detection of mature miRNAs/small RNAs, we followed the protocol described in ref. 26, with some modifications. Briefly, total RNA was purified using Trizol following the manufacturer's instruction; 5 μg of total RNA was polyadenylated with *Escherichia coli* PAP polymerase (Ambion). First strand cDNA was generated using a polyT adaptor 5′-GCGAGCACAGAATTAATACGACTCACTATAGGTTTTTTTTTTTTVN-3′. PCR products corresponding to mature miRNA were amplified (36–39 cycles) using a universal reverse primer and a forward primer specific for each mature miRNA (Supplementary Table 2).

The Methods section describes how DNA constructs were generated and introduced in cells and embryos; biological and biochemical assays in embryos and human cells; and *in situ* hybridization and immunohistochemistry.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 February; accepted 18 July 2007.
Published online 29 August 2007.

1. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116,** 281–297 (2004).
2. Ambros, V. The functions of animal microRNAs. *Nature* **431,** 350–355 (2004).
3. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
4. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).
5. Massague, J. How cells read TGF-β signals. *Nature Rev. Mol. Cell Biol.* **1,** 169–178 (2000).
6. Niehrs, C. Regionally specific induction by the Spemann-Mangold organizer. *Nature Rev. Genet.* **5,** 425–434 (2004).
7. Schohl, A. & Fagotto, F. β-catenin, MAPK and Smad signaling during early *Xenopus* development. *Development* **129,** 37–52 (2002).
8. Faure, S., Lee, M. A., Keller, T., ten Dijke, P. & Whitman, M. Endogenous patterns of TGFβ superfamily signaling during early *Xenopus* development. *Development* **127,** 2917–2931 (2000).
9. De Robertis, E. M., Larrain, J., Oelgeschlager, M. & Wessely, O. The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nature Rev. Genet.* **1,** 171–181 (2000).
10. Tao, Q. *et al.* Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in *Xenopus* embryos. *Cell* **120,** 857–871 (2005).
11. Heasman, J., Kofron, M. & Wylie, C. β-catenin signaling activity dissected in the early *Xenopus* embryo: a novel antisense approach. *Dev. Biol.* **222,** 124–134 (2000).
12. Heasman, J. *et al.* Overexpression of cadherins and underexpression of β-catenin inhibit dorsal mesoderm induction in early *Xenopus* embryos. *Cell* **79,** 791–803 (1994).
13. Agius, E., Oelgeschlager, M., Wessely, O., Kemp, C. & De Robertis, E. M. Endodermal Nodal-related signals and mesoderm induction in *Xenopus*. *Development* **127,** 1173–1183 (2000).
14. Pogoda, H. M., Solnica-Krezel, L., Driever, W. & Meyer, D. The zebrafish forkhead transcription factor FoxH1/Fast1 is a modulator of nodal signaling required for organizer formation. *Curr. Biol.* **10,** 1041–1049 (2000).

15. Cha, Y. R., Takahashi, S. & Wright, C. V. Cooperative non-cell and cell autonomous regulation of Nodal gene expression and signaling by Lefty/Antivin and Brachyury in *Xenopus*. *Dev. Biol.* **290,** 246–264 (2006).
16. Norris, D. P. & Robertson, E. J. Asymmetric and node-specific nodal expression patterns are controlled by two distinct cis-acting regulatory elements. *Genes Dev.* **13,** 1575–1588 (1999).
17. Vize, P. D. DNA sequences mediating the transcriptional response of the Mix.2 homeobox gene to mesoderm induction. *Dev. Biol.* **177,** 226–231 (1996).
18. Chen, X. *et al.* Smad4 and FAST-1 in the assembly of activin-responsive factor. *Nature* **389,** 85–89 (1997).
19. Krek, A. *et al.* Combinatorial microRNA target predictions. *Nature Genet.* **37,** 495–500 (2005).
20. John, B. *et al.* Human MicroRNA targets. *PLoS Biol.* **2,** e363 (2004).
21. Song, J. *et al.* The type II activin receptors are essential for egg cylinder growth, gastrulation, and rostral head development in mice. *Dev. Biol.* **213,** 157–169 (1999).
22. Watanabe, T. *et al.* Stage-specific expression of microRNAs during *Xenopus* development. *FEBS Lett.* **579,** 318–324 (2005).
23. Krutzfeldt, J., Poy, M. N. & Stoffel, M. Strategies to determine the biological function of microRNAs. *Nature Genet.* **38** (Suppl), S14–S19 (2006).
24. Flynt, A. S., Li, N., Thatcher, E. J., Solnica-Krezel, L. & Patton, J. G. Zebrafish miR-214 modulates Hedgehog signaling to specify muscle cell fate. *Nature Genet.* **39,** 259–263 (2007).
25. Piccolo, S. *et al.* The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals. *Nature* **397,** 707–710 (1999).
26. Shi, R. & Chiang, V. L. Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* **39,** 519–525 (2005).
27. Feldman, B. *et al.* Lefty antagonism of Squint is essential for normal gastrulation. *Curr. Biol.* **12,** 2129–2135 (2002).
28. Meno, C. *et al.* Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4,** 287–298 (1999).
29. Zhang, L. *et al.* Zebrafish Dpr2 inhibits mesoderm induction by promoting degradation of nodal receptors. *Science* **306,** 114–117 (2004).
30. Giraldez, A. J. *et al.* MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308,** 833–838 (2005).
31. Bernstein, E. *et al.* Dicer is essential for mouse development. *Nature Genet.* **35,** 215–217 (2003).
32. Tang, F. *et al.* Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.* **21,** 644–648 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Bressan and D. Volpin for discussion. This work is supported by grants from AIRC, TELETHON-Italy, MIUR (CoFin, FIRB), ASI, the ISS-Stem cells program and Swissbridge to S.P. A.M. is a recipient of an EU-Marie Curie RTN fellowship (epiplast carcinoma). We are grateful to J. Moulton for help in the design of miRNA morpholinos; C. Niehrs, N. Ueno, J. Green, W. Knochel and M. Asashima for gifts of plasmids; and W. Vale for the anti-Acvr2a antibody and F. Fagotto for protocols. L.Z. is a recipient of a post-doctoral contract from the University of Padua and M.I. is a recipient of a TOYOBO Biotechnology Foundation (Japan) grant.

Author Contributions G.M. identified Acvr2a as a target of miR-15 and miR-16. G.M., L.Z. and M.I. performed the *Xenopus* assays. M.C. and G.M. carried out experiments in human cells. U.T. and L.Z. performed the immunohistochemistry analysis. S.P. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.P. (piccolo@bio.unipd.it).

METHODS

Plasmid and oligonucleotides. Endogenous miRNAs were purified as previously described[22]. For the generation of miR-15 reporter constructs (Acvr2a-WT and Acvr2a-mut), the luciferase cDNA was first inserted in pcDNA3.1; the first miR-15 and miR-16 binding site of *Acvr2a* 3′ UTR was then cloned downstream of luciferase by inserting double-stranded oligonucleotides. Mature miR-15 sequences were: wild-type sense, 5′-UAGCAGCACAUAAUGGUUUGUGUU-3′; wild-type antisense, 5'CACAAACCAUUAUGUGCUGGAUUU-3′; mutant sense, 5′-UUCGUCAACAUAAUGGUUUGUGUU-3′; mutant antisense, 5′-CACAAACCAUUAUGUUGACCUUUU-3′ (Invitrogen). Anti-miRNA reagents were purchased from Dharmacon: anti-miR-15, 5′-AAUCCACAAACCAUU-AUGUGCUGCUACUUU-3′; anti-miR-16, 5′-AAUCCUCCAGUAUUUACGU-GCUGCUAAGGC-3′; control anti-miR-16 sense, 5′- UAGCAGCACGUA-AAUACUGGAG -3′; and control scrambled anti-miRNA. Pri-miR-15/16a and pri-miR-127 were amplified from genomic DNA by PCR and cloned into pCS2 and sequenced. RNA was synthesized using Message Machine (Ambion). Of note, we found that *in vitro* transcribed miRNA was biologically effective *in vivo* only if capped and containing the SV40 polyadenylation signal. Moreover, in order to facilitate the formation of proper secondary structures (hairpins), pri-miRNAs were denatured at 65 °C for 5 min and slowly annealed to RT.

miR-15 and miR-16 morpholino oligonucleotides (Gene Tools) were 25-base oligonucleotides designed to be complementary to the mature miRNA, but extending one or two bases over the flanking sequence of guide strand, past the Drosha and Dicer cleavage sites: *mir-15b* MO, 5′-CATGCAAATCATGA-TGTGCTGCTAC-3′; *mir-16b* MO, 5′-TCACCCAATATTTACGTGCTGCTAA-3′; *mir-15c* MO, 5′-TCTACAAACCATGATGTGCTGCTAG-3′; *mir-16c* MO, 5′-AACTCCAGTATTTACGTGCTGCTAA-3′; control MO, 5′-CCTCTTACC-TCAGTTACAATTTATA-3′.

Acvr2a silencer validated siRNA was purchased from Ambion: sense, 5′-GGACUGAUUGUGUAGAAAAtt-3′; antisense, 5′-UUUUCUACACAAUCA-GUCCtg-3′.

Biological and biochemical assays. *Xenopus* embryo manipulations, mRNA preparation, microinjections and whole-mount *in situ* hybridizations were performed as previously described[25]. Mature miRNAs, anti-miRNAs and morpholino oligonucleotides were re-suspended in 0.5 mM HEPES, pH 7.6. For luciferase assays, embryos were radially injected (except in Fig. 4c) with 40 pg of reporter plasmid (UTR, Mix.2 or Vent.2) plus 150 pg of *lacZ* mRNA and collected at early gastrula. Luciferase values were normalized to β-gal activity that, typically, did not oscillate more that 25% within the same batch. For each experiment, the normalized luciferase value is the mean of at least three independent embryo sets, each containing five embryos. In Fig. 5b we measured the effects of Wnt/β-catenin on endogenous miR-15 and miR-16 considering the fold inhibition (typically 50%, as in Fig. 1b) between Acvr2a-mut and wild-type reporter. HepG2 cells were transfected with Lipofectamine2000 (Invitrogen) in DMEM, 10% serum with reporter plasmids (40 ng cm^{-2}) and miR-15 RNA (300 ng cm^{-2}); pCS2-lacZ (20 ng cm^{-2}) was used as a normalizer. For experiments shown in Fig. 1d, HepG2 cells were first transfected with *Acvr2a* siRNA or control siRNA (siGFP) using RNAi max (Invitrogen); after 48 h, cells were transfected with reporter plasmids (40 ng cm^{-2}) and pCS2-lacZ (20 ng cm^{-2}) together with the indicated miR-15 RNA (300 ng cm^{-2}) and pCS2Acvr2a-3′UTR deleted (400 ng cm^{-2}). A detailed protocol has been uploaded to Protocols Network (doi:10.1038/nprot.2007.349). In each experiment, samples were transfected in duplicate. Treatments of cells with activinA, TGF-β1 and Bmp2 proteins were carried out in 0.1% serum.

Western blotting was carried out as described[33]. Anti-Acvr2a-specific polyclonal antibody was purchased from R&D. Quantifications were performed using NIH-Image software.

Immunohistochemistry. For immunostaining, embryos were fixed for 2 h at room temperature in MEMFA then in 80% methanol/20% dimethyl sulphoxide overnight at −20 °C. They were rinsed in PBS, dehydrated, embedded in paraffin and cut into 20-μm sections. Slides were de-waxed, re-hydrated and treated for 45 min with 2% H_2O_2 to quench the endogenous peroxidase activity. Antigen retrieval was performed by microwave boiling in citrate buffer, pH 6.0. Slides were then blocked for 30 min in PBS, 0.1% Tween-20, 10% goat serum. Slides were then incubated overnight at 4 °C with the anti-Acvr2a antibody (1:250, gift of W. Vale) in blocking buffer. We used the Vectastain ABC-Elite kit (Vector Laboratories) to reveal the signal according to the manufacturer's instruction. The slides were dehydrated and mounted for microscopy. The sections were observed under a Zeiss Axioplan microscope equipped with a Leica DC500 camera.

33. Cordenonsi, M. *et al.* Integration of TGF-β and Ras/MAPK signaling through p53 phosphorylation. *Science* **315**, 840–843 (2007).

LETTERS

A giant planet orbiting the 'extreme horizontal branch' star V 391 Pegasi

R. Silvotti[1], S. Schuh[2], R. Janulis[3], J.-E. Solheim[4], S. Bernabei[5], R. Østensen[6], T. D. Oswalt[7], I. Bruni[5], R. Gualandi[5], A. Bonanno[8], G. Vauclair[9], M. Reed[10], C.-W. Chen[11], E. Leibowitz[12], M. Paparo[13], A. Baran[14], S. Charpinet[9], N. Dolez[9], S. Kawaler[15], D. Kurtz[16], P. Moskalik[17], R. Riddle[18] & S. Zola[14,19]

After the initial discoveries fifteen years ago[1,2], over 200 extrasolar planets have now been detected. Most of them orbit main-sequence stars similar to our Sun, although a few planets orbiting red giant stars have been recently found[3]. When the hydrogen in their cores runs out, main-sequence stars undergo an expansion into red-giant stars. This expansion can modify the orbits of planets and can easily reach and engulf the inner planets. The same will happen to the planets of our Solar System in about five billion years and the fate of the Earth is matter of debate[4,5]. Here we report the discovery of a planetary-mass body ($M\sin i = 3.2 M_{\rm Jupiter}$) orbiting the star V 391 Pegasi at a distance of about 1.7 astronomical units (AU), with a period of 3.2 years. This star is on the extreme horizontal branch of the Hertzsprung–Russell diagram, burning helium in its core and pulsating. The maximum radius of the red-giant precursor of V 391 Pegasi may have reached 0.7 AU, while the orbital distance of the planet during the stellar main-sequence phase is estimated to be about 1 AU. This detection of a planet orbiting a post-red-giant star demonstrates that planets with orbital distances of less than 2 AU can survive the red-giant expansion of their parent stars.

With an effective temperature close to 30,000 K and a surface gravity ten times that of the Sun[6], V 391 Pegasi (or HS 2201+2610 from the original Hamburg Schmidt survey name) is one of about 40 hot subdwarf B stars showing short-period p-mode pulsations[7]. Its pulsational spectrum exhibits four or five pulsation periods[6,8] between 342 and 354 s (see Supplementary Information for more details on the star's properties).

Because of their compact structure, subdwarf B pulsators have extremely stable oscillation periods, like white dwarf pulsators. It is therefore possible to register very small differences in the arrival times of the photons[9,10], which in principle allows the detection of low-mass secondary bodies[11], through the use of the observed–calculated (O–C) diagram[12] (see Fig. 1 legend for more details). Functionally, it is equivalent to the timing method used to find planets around pulsars[1,13].

When a pulsation period changes linearly in time, the O–C diagram has a parabolic shape, as confirmed by all the previous measurements of $\mathrm{d}P/\mathrm{d}t$ (or $\dot{P}$) in compact pulsators[9,14,15]. The same behaviour is found in the O–C plot of V 391 Peg (upper panel of Fig. 1), implying that the main pulsation period of the star is increasing at a rate of $\dot{P}_1 = (1.46 \pm 0.07) \times 10^{-12}$ (or 1 s in 22,000 years).

For V 391 Peg a simple second-order polynomial does not give a satisfactory fit and the sinusoidal residuals require further interpretation (lower panel of Fig. 1). An oscillating $\dot{P}$ is not compatible with any evolutionary or pulsational model: it would require that the star expands and contracts periodically every 3 years, a time much larger than the dynamical timescale, which is of the order of 500 s for a subdwarf B star. Nor can the sinusoidal residuals be explained by any known pulsational effect. Random period variations[16] are also not a possibility because these variations would be cancelled by the large number of pulsation cycles (>28 for each point in Fig. 1).

The simplest explanation for the sinusoidal component of the O–C diagram in Fig. 1 is a wobble of the star's barycentre due to the presence of a low-mass companion. Depending on its position around the barycentre of the system, the subdwarf B star is periodically closer to, or more distant from us by 5.3 ± 0.6 light seconds and the timing of the pulsation is cyclically advanced or delayed. From our best fit and Kepler's third law (assuming a circular orbit, $M_1 = 0.5 M_{\rm Sun}$, where $M_{\rm Sun}$ is the mass of the Sun, and $M_2 \ll M_1$), we obtain: $P_{\rm orb} = 1{,}170 \pm 44$ days (or 3.20 ± 0.12 years), $a = 1.7$ AU and $M_2 \sin i = 3.2 M_{\rm Jupiter}$, where a is the planet–star separation and 1 AU is the mean distance between the Earth and the Sun. The orbital parameters of the system are listed in Table 1. This interpretation is robust: the

Table 1 | Orbital parameters

Parameter	Value
Orbital period, $P_{\rm orb}$ (d)	$1{,}170 \pm 44$
Epoch of maximum time delay, T_0 (BJD)	$2{,}452{,}418 \pm 96$
Eccentricity, e (assumed)	0.0
Star projected orbital radius, $a_s \sin i$ (km)	$1{,}600{,}000 \pm 190{,}000$
Star projected orbital velocity, $v_s \sin i$ ($\mathrm{m\,s^{-1}}$)	99 ± 12
Mass function*, $f(M_1, M_2)$ ($M_{\rm Sun}$)	$(1.19 \pm 0.43) \times 10^{-7}$
Distance from the star†, a (AU)	1.7 ± 0.1
Maximum elongation† (milliarcsec)	1.2 ± 0.1
Planet orbital velocity†, v_p ($\mathrm{km\,s^{-1}}$)	16 ± 1
Planet mass†, $M_2 \sin i$ ($M_{\rm Jupiter}$)	3.2 ± 0.7

* $f(M_1,M_2) = 4\pi^2 (a_s \sin i)^3 \big/ G P_{\rm orb}^2 = (M_2 \sin i)^3 \big/ (M_1 + M_2)^2$.

† These numbers are obtained assuming $M_1 = 0.5 \pm 0.05 M_{\rm Sun}$ (suggested from asteroseismology) and $M_2 \ll M_1$.

[1]INAF-Osservatorio Astronomico di Capodimonte, via Moiariello 16, 80131 Napoli, Italy. [2]Institut für Astrophysik, Universität Göttingen, Friedrich-Hund-Platz 1, 37077 Göttingen, Germany. [3]Institute of Theoretical Physics and Astronomy, Vilnius University, 12 A. Gostauto Street, 01108 Vilnius, Lithuania. [4]Institutt for Teoretisk Astrofysikk, Universitetet i Oslo, PB 1029 Blindern, 0315, Norway. [5]INAF-Osservatorio Astronomico di Bologna, via Ranzani 1, 40127 Bologna, Italy. [6]K. U. Leuven, Institute of Astronomy, Celestijnenlaan 200D, 3001 Leuven, Belgium. [7]Department of Physics and Space Sciences and the SARA Observatory, Florida Institute of Technology, 150 West University Boulevard, Melbourne, Florida 32901, USA. [8]INAF-Osservatorio Astrofisico di Catania, via S. Sofia 78, 95123 Catania, Italy. [9]CNRS-UMR5572, Observatoire Midi-Pyrénées, Université Paul Sabatier, 14 avenue Edouard Belin, 31400 Toulouse, France. [10]Department of Physics, Astronomy and Materials Science, Missouri State University, 901 S. National, Springfield, Missouri 65897, USA. [11]Institute of Astronomy, National Central University, 300 Jhongda Road, Chung-Li 32054, Taiwan. [12]Wise Observatory, Tel Aviv University, Tel Aviv 69978, Israel. [13]Konkoly Observatory, P O Box 67, H-1525 Budapest XII, Hungary. [14]Cracow Pedagogical University, ul. Podchorazych 2, 30-084 Cracow, Poland. [15]Department of Physics and Astronomy, 12 Physics Hall, Iowa State University, Ames, Iowa 50011, USA. [16]Centre for Astrophysics, University of Central Lancashire, Preston PR1 2HE, UK. [17]Copernicus Astronomical Centre, ul. Bartycka 18, 00-716 Warsaw, Poland. [18]Thirty Meter Telescope Project, 2632 E. Washington Blvd, Pasadena, California 91107, USA. [19]Astronomical Observatory, Jagiellonian University, ul. Orla 171, 30-244 Cracow, Poland.

same sinusoidal component is also found in the O–C diagram of the secondary pulsation frequency of the star (see Figs 2 and 3 for more details). Any alternative interpretation of our results would have to be compatible with this fact. The sinusoids in the lower panels of Figs 1, 2 and 3 suggest a circular orbit. From our observations we cannot yet set a precise upper limit to the eccentricity, but it must be close to zero.

Using the known characteristics of the V 391 Peg system, we can determine a first estimate of the planet's effective temperature by balancing the flux received from the star with the blackbody flux re-radiated by the planet (see Supplementary Information for more details). Assuming a Bond albedo of 0.343 (similar to that of Jupiter[17]), we obtain an effective temperature for the planet of about 470 K, corresponding to a maximum of the blackbody radiation near 6.2 μm from Wien's law.

With a projected radius of about five light seconds, the wobble of the barycentre of V 391 Peg points towards a planet (for comparison, the amplitude of the solar displacement around the barycentre of our Solar System is almost three light seconds). However, depending on the unknown inclination i of the system, a brown dwarf or even a low-mass stellar companion cannot be totally excluded. But the low inclination required ($2.5^\circ \leq i \leq 14^\circ$ for a brown dwarf or $i \leq 2.5^\circ$ for a low-mass stellar companion) has a very low probability (3% and 0.1% respectively), assuming a random distribution of orbital plane inclinations.

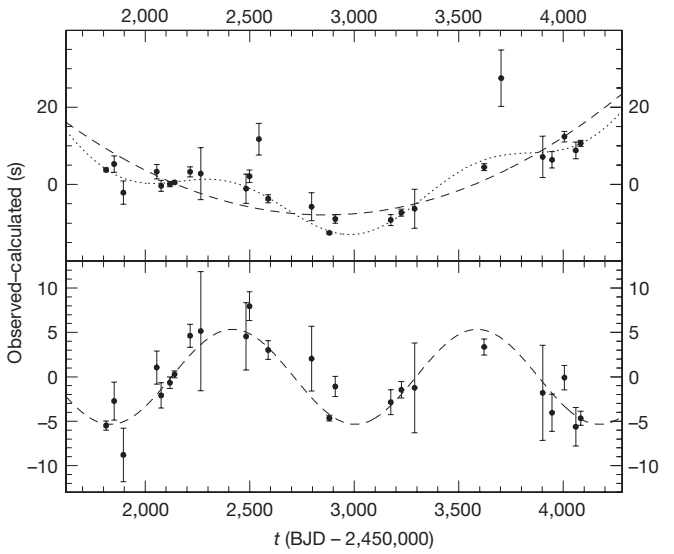


Figure 1 | The O–C diagram of the main pulsation frequency f_1 of V 391 Peg. The O–C technique is a way of measuring the phase variations of a periodic function, comparing the observed times of the maxima with those calculated from an ephemeris[12]. In our case, what is compared is the time of the first maximum of each single run (obtained by fitting the data with five sinusoids simultaneously, corresponding to the five pulsation frequencies) with the best ephemeris obtained fitting the whole (seven-year-long) data set. The error bars are given by $\left(\sigma_O^2+\sigma_C^2\right)^{1/2}$, where σ_O and σ_C are the 1σ phase errors obtained from the least-squares sinusoidal fits. The upper panel shows that the fit of the long-term component by a second-order polynomial is significantly improved when we also use a sine wave. Fitting the data with both functions simultaneously reduces the value of the reduced χ^2 from 14.1 (second-order polynomial alone) to 2.7. The lower panel shows the sinusoidal component alone. To obtain these plots, 418 h of time-series photometry from 167 nights of observation were used, from 18 different 1-m to 3-m class telescopes (see Supplementary Information for more details). The number of photometric measurements for each point varies from 237 (the highest point with large error bar) to 26,081 (the first point on the left). In total, the number of photometric measurements is 109,531. The data were reduced following standard procedures for time-series photometry, using statistical weights[27] and barycentric time corrections[28] (BJD stands for barycentric Julian day). One second was added to the data of 2006 only, to compensate for the leap-second correction of 1 January 2006. Looking at the time distribution of the phase measurements, we note that there are seven groups of close points corresponding to the seven observing seasons (from May to December) of the last seven years (2000 to 2006).

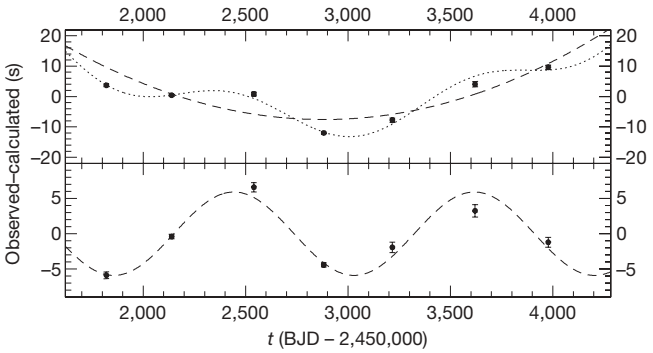


Figure 2 | The O–C diagram of f_1. In this version of the O–C diagram, all the runs of each observing season were combined and the phases were recalculated on these larger data sets. This reduces the noise (but also reduces the time resolution), so that in principle O–C diagrams can be built for each pulsation frequency of a multiperiodic pulsator. In this way, if the pulsating star has a companion, each pulsation mode can supply an independent confirmation of the periodic motion around the centre of mass. V 391 Peg has four or five pulsation periods; for the two that have sufficiently large amplitudes of 1% and 0.4% respectively, O–C diagrams can be obtained. As in Fig. 1, the upper and lower plots represent respectively the O–C diagram of f_1 and its sinusoidal component alone. The error bars are calculated as in Fig. 1.

Thus, with a 97% probability, V 391 Peg b is the first recognized planet orbiting a post-red-giant star, making this system a unique laboratory in which to test the evolution of planetary systems during and after the red-giant expansion. With a probable age of the order of 10 Gyr (see Supplementary Information for more details), V 391 Peg b is also one of the oldest planets known. An interesting case of a brown dwarf that survived engulfment by a red giant was recently presented[18]; the information about whether low-mass companions to red-giant stars survive engulfment in that system is complementary to that of V 391 Peg, because the two systems are very different. We note that only by studying planets of horizontal branch stars is it possible to

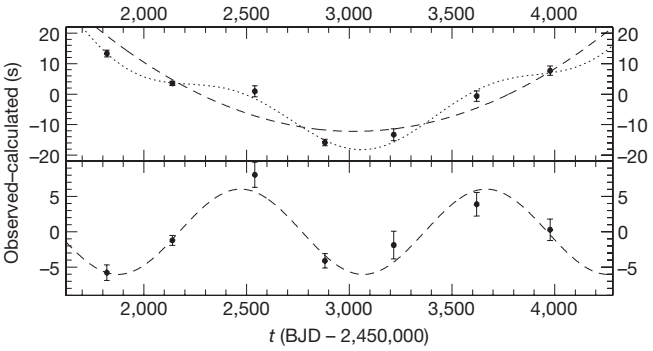


Figure 3 | The O–C diagram of f_2. As for Fig. 2 but relative to the second pulsation frequency f_2. Comparing the lower panels of Figs 2 and 3, we see that the two sinusoids of f_1 and f_2 are identical within the errors. The agreement between periods, amplitudes and phases is always better than 0.2σ. We obtain respectively 1,174 ± 94 days versus 1,194 ± 106 days, 5.9 ± 1.6 s versus 6.0 ± 2.3 s, and BJD 2,452,443 ± 194 versus BJD 2,452,471 ± 211 for the epoch of the first maximum. From the second-order polynomial component of the fit in the upper panel, we obtain also a measurement of the secular variation of f_2: $\dot{P}_2 = (2.05 \pm 0.26) \times 10^{-12}$, which is different from $\dot{P}_1 = (1.46 \pm 0.07) \times 10^{-12}$. The absolute values of $\dot{P}_1$ and $\dot{P}_2$, which correspond to an evolutionary timescale $P/\dot{P}$ of 7.6×10^6 and 5.5×10^6 years respectively, match relatively well with theoretical expectations for evolved models of extreme horizontal branch stars[29] (even though their positive sign is more difficult to explain and suggests that the star is expanding, as confirmed by some tests done by one of us). We note that the difference between $\dot{P}_1$ and $\dot{P}_2$ excludes the possibility that the long-term component of the O–C plots is due to a secondary planet with a larger orbit. The error bars are calculated as in Fig. 1.

isolate the effects of the red-giant expansion on a planetary system. Planets around white dwarfs must be strongly modified also by the asymptotic giant branch expansion, the thermal pulses and the final planetary nebula ejection[19].

Even though, in terms of orbit stability, the existence of V 391 Peg b is not surprising[20], in terms of orbit evolution during the red-giant phase, the situation is less clear. There are at least two competing processes that determine the orbital evolution: mass loss from the star that causes the orbit of a planet to expand, and tidal effects that tend to reduce its angular momentum causing spiralling-in[21]. Neither the stellar mass loss nor the tidal dissipation are well-understood processes. For this reason, the destiny of our Earth is still a matter of debate[4,5]. For V 391 Peg b the most likely scenario is that the planet never entered the stellar envelope (the maximum radius expected for a subdwarf B progenitor at the tip of the red-giant branch[22,23] is of the order of 0.7 AU) and that the orbit of V 391 Peg b was tighter in the past owing to the strong mass loss of the parent star, with an orbital radius of about 1 AU when the star was still on the main sequence. This value is obtained by assuming that the stellar mass has been reduced from $0.85M_{\rm Sun}$ to $0.5M_{\rm Sun}$, when tidal interaction (which is proportional to $(R_*/r)^8$; ref. 24) can be neglected for a sufficiently large orbital distance r with respect to the stellar radius R_*. In this scenario the increase of the planet's mass due to accretion from the stellar wind is negligible[20]. We note that in this case, incidentally, the orbital distances of V 391 Peg b and of the Earth, before and after the red-giant phase, are very similar: 1.5 AU is a reasonable value for the Earth after red-giant migration, when tidal effects are not considered[4,5].

A different scenario is obtained if the mass loss of the red-giant precursor of V 391 Peg started sufficiently late: in this case the ratio between stellar radius and orbital distance could have reached a value of about 0.7, at which the star fills its Roche lobe[25] and mass transfer to the planet starts, causing the planet to spiral quickly into the outer layers of the giant's atmosphere. Here accretion is disrupted and the spiral-in due to accretion stops, so that the planet may have survived if the spiral-in due to friction was sufficiently low. The presence of planets with orbital separations $\lesssim 5$ AU has been invoked by a few authors to explain the strong mass loss needed to form subdwarf B stars and partially explain the irregular morphology of the horizontal branch[26].

Received 6 April; accepted 26 July 2007.

1. Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR1257+12. *Nature* **355,** 145–147 (1992).
2. Mayor, M. & Queloz, D. A Jupiter-mass companion to a solar-type star. *Nature* **378,** 355–359 (1995).
3. Döllinger, M. P. *et al.* Discovery of a planet around the K giant star 4 U Ma. *Astron. Astrophys.* (in the press); preprint at ⟨http://arxiv.org/astro-ph/0703672⟩.
4. Rasio, F. A., Tout, C. A., Lubow, S. H. & Livio, M. Tidal decay of close planetary orbits. *Astrophys. J.* **470,** 1187–1191 (1996).
5. Rybicki, K. R. & Denis, C. On the final destiny of the Earth and the Solar System. *Icarus* **151,** 130–137 (2001).
6. Østensen, R. *et al.* Detection of pulsations in three subdwarf B stars. *Astron. Astrophys.* **368,** 175–182 (2001).
7. Kilkenny, D. Pulsating hot subdwarfs—an observational review. *Commun. Asteroseismol.* **150,** 234–240 (2007).
8. Silvotti, R. *et al.* The temporal spectrum of the sdB pulsating star HS 2201+2610 at 2 ms resolution. *Astron. Astrophys.* **389,** 180–190 (2002).
9. Kepler, S. O. *et al.* Measuring the evolution of the most stable optical clock G 117–B15A. *Astrophys. J.* **634,** 1311–1318 (2005).
10. Reed, M. *et al.* Observations of the pulsating subdwarf B star Feige 48: constraints on evolution and companions. *Mon. Not. R. Astron. Soc.* **348,** 1164–1174 (2004).
11. Winget, D. E. *et al.* The search for planets around pulsating white dwarf stars. *ASP Conf. Ser.* **294,** 59–64 (2003).
12. Sterken, C. The O–C diagram: basic procedures. *ASP Conf. Ser.* **335,** 3–23 (2005).
13. Thorsett, S. E., Arzoumanian, Z. & Taylor, J. H. PSR B1620–26—A binary radio pulsar with a planetary companion? *Astrophys. J.* **412,** L33–L36 (1993).
14. Costa, J. E. S., Kepler, S. O. & Winget, D. E. Direct measurement of a secular pulsation period change in the pulsating hot pre-white dwarf PG 1159–035. *Astrophys. J.* **522,** 973–982 (1999).
15. Mukadam, A. S. *et al.* Constraining the evolution of ZZ Ceti. *Astrophys. J.* **594,** 961–970 (2003).
16. Koen, C. Statistics of O–C diagrams and period changes. *ASP Conf. Ser.* **335,** 25–35 (2005).
17. Hanel, R., Conrath, B., Herath, L., Kunde, V. & Pirraglia, J. Albedo, internal heat, and energy balance of Jupiter—preliminary results of the Voyager infrared investigation. *J. Geophys. Res.* **86,** 8705–8712 (1981).
18. Maxted, P. F. L., Napiwotzki, R., Dobbie, P. D. & Burleigh, M. R. Survival of a brown dwarf after engulfment by a red giant star. *Nature* **442,** 543–545 (2006).
19. Villaver, E. & Livio, M. Can planets survive stellar evolution? *Astrophys. J.* **661,** 1192–1201 (2007).
20. Duncan, M. J. & Lissauer, J. J. The effects of post-main-sequence solar mass loss on the stability of our planetary system. *Icarus* **134,** 303–310 (1998).
21. Livio, M. & Soker, N. Star-planet systems as progenitors of cataclysmic binaries: tidal effects. *Astron. Astrophys.* **125,** L12–L15 (1983).
22. Sweigart, A. V. & Gross, P. G. Evolutionary sequences for red giant stars. *Astrophys. J.* **36** (Suppl.), 405–437 (1978).
23. Han, Z., Podsiadlowski, Ph, Maxted, P. F. L., Marsh, T. R. & Ivanova, N. The origin of subdwarf B stars—I. The formation channels. *Mon. Not. R. Astron. Soc.* **336,** 449–466 (2002).
24. Zahn, J. P. Tidal friction in close binary stars. *Astron. Astrophys.* **57,** 383–394 (1977).
25. Eggleton, P. P. Approximations to the radii of Roche lobes. *Astrophys. J.* **268,** 368–369 (1983).
26. Soker, N. Can planets influence the horizontal branch morphology? *Astron. J.* **116,** 1308–1313 (1998).
27. Silvotti, R. *et al.* The rapidly pulsating subdwarf B star PG 1325+101. I. Oscillation modes from multisite observations. *Astron. Astrophys.* **459,** 557–564 (2006).
28. Stumpff, P. Two self-consistent FORTRAN subroutines for the computation of the Earth's motion. *Astron. Astrophys. Suppl. Ser.* **41,** 1–8 (1980).
29. Charpinet, S., Fontaine, G., Brassard, P. & Dorman, B. Adiabatic survey of subdwarf B star oscillations. III. Effects of extreme horizontal branch stellar evolution on pulsation modes. *Astrophys. J.* **140** (Suppl.), 469–561 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements R.S. thanks M. Capaccioli, J. M. Alcalá, E. Covino and S. O. Kepler for discussions and suggestions, S. Marinoni and S. Galleti for their contribution to the observations, and the MiUR for financial support. S.S. thanks T. Nagel, E. Goehler, T. Stahn, S. D. Huegelmeyer, R. Lutz, U. Thiele and A. Guijarro for their help in data acquisition, and the DFG for travel grants. R.Ø. is supported by the Research Council of the University of Leuven and by the FP6 Coordination Action HELAS of the EU. T.D.O. acknowledges support from the US National Science Foundation. P.M. acknowledges support from the Polish MNiSW.

Author Contributions R.S. analysed and interpreted the data from which the presence of the planet was inferred. R.S., S.S., R.J., J.-E.S., S.B., R.Ø., T.D.O., I.B., R.G., A. Bonanno., G.V., M.R., C.-W.C., E.L. and M.P. contributed to the large amount of observations and/or data reduction. A. Baran., S.C., N.D., S.K., D.K., P.M., R.R. and S.Z. contributed to the organization and/or on-line data reduction/analysis during the XCov23 Whole Earth Telescope campaign of August-September 2003, in which V 391 Peg was observed as a secondary target. S.K. performed some tests on theoretical $\dot{P}$. S.K. and S.Z. did independent checks of the O–C fits. E.L. made statistical tests on the significance level of the O–C fits. All authors discussed and interpreted the results and commented on the manuscript. D.K. and R.Ø. in particular helped to improve the text.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.S. (silvotti@na.astro.it).

LETTERS

Dynamics of ice ages on Mars

Norbert Schorghofer[1]

Unlike Earth, where astronomical climate forcing is comparatively small, Mars experiences dramatic changes in incident sunlight that are capable of redistributing ice on a global scale[1–6]. The geographic extent of the subsurface ice found poleward of approximately ±60° latitude on both hemispheres of Mars[7–9] coincides with the areas where ice is stable[7,10,11]. However, the tilt of Mars' rotation axis (obliquity) changed considerably in the past several million years. Earlier work[3,12] has shown that regions of ice stability, which are defined by temperature and atmospheric humidity, differed in the recent past from today's, and subsurface ice is expected to retreat quickly when unstable[11–13]. Here I explain how the subsurface ice sheets could have evolved to the state in which we see them today. Simulations of the retreat and growth of ground ice as a result of sublimation loss and recharge reveal forty major ice ages over the past five million years. Today, this gives rise to pore ice at mid-latitudes and a three-layered depth distribution in the high latitudes of, from top to bottom, a dry layer, pore ice, and a massive ice sheet. Combined, these layers provide enough ice to be compatible with existing neutron and gamma-ray measurements[7–9].

The extensive subsurface ice deposits on Mars[7–9] were anticipated on the basis of atmosphere–subsurface vapour exchange[3,12,14]. Entirely unexpected, however, was the large amount of ground ice, 60% by weight or 70–85% by volume in the near-surface of the planet[15], which is far greater than the porosity of typical soils. There are two possible mechanisms for the emplacement of the ice: (1) precipitation from the atmosphere[2,4,5] or (2) diffusion and condensation of atmospheric vapour in regolith pores[12]. The high ice content suggests that the former mechanism was at work. Indeed, climate simulations show that at high obliquity water ice from the north polar cap precipitates in the tropics[1,2], and that as the obliquity decreases, the low-latitude ice is redistributed to form the mid-latitude ice sheets[4,5]. Subsequently, the ice sheets evolved; today the ice is buried and in its equilibrium position, where the vapour pressure of the ice balances the atmospheric humidity[7,10,11]. When an ice sheet retreats, the dust it contains can form a lag deposit that protects the ice from rapid sublimation and retreat. But still, the time over which the subsurface ice is expected to adjust to drier or warmer conditions is comparable to that over which changes in Mars' obliquity and orbital parameters occur[11–13]. Theoretical as well as experimentally measured diffusion coefficients of water vapour through soils[13] are orders of magnitude too large to isolate ice from the atmosphere over millions of years.

Although there are two mechanisms for ice accumulation, the rate-limiting process for the loss of buried ice is always diffusion. A model of subsurface temperature and ice stability is designed (and described in the Methods). It is initiated with a cover of dirty ice, consisting of 85% ice and 15% dust. Subsequently, ice is lost by diffusion through the sublimation lag, which is assigned a realistic diffusivity[13] of 4 $cm^2\,s^{-1}$. Ice can re-form by inward diffusion of atmospheric water, to fill the interstitial pores with a porosity of 40%. These climate simulations are computationally affordable because of a new numerical approach that uses averaging methods for the diffusion equation, rendered nonlinear by phase transitions. Such averaging methods have been applied to the retreat of an ice table[11,16] and can be extended to partially filled pores. The climate model does not keep track of the global H_2O balance, and every latitude is independently computed on a distributed computer cluster. It is assumed a global ice sheet formed by atmospheric precipitation five million years ago[5], and for simplicity only a single ice sheet is considered. This climate model integrates history using time-varying orbital elements (ref. 17 and see its data at www.imcce.fr/Equipes/ASD/insola/mars/mars.html). Zonally averaged values of present-day albedo and thermal properties are used.

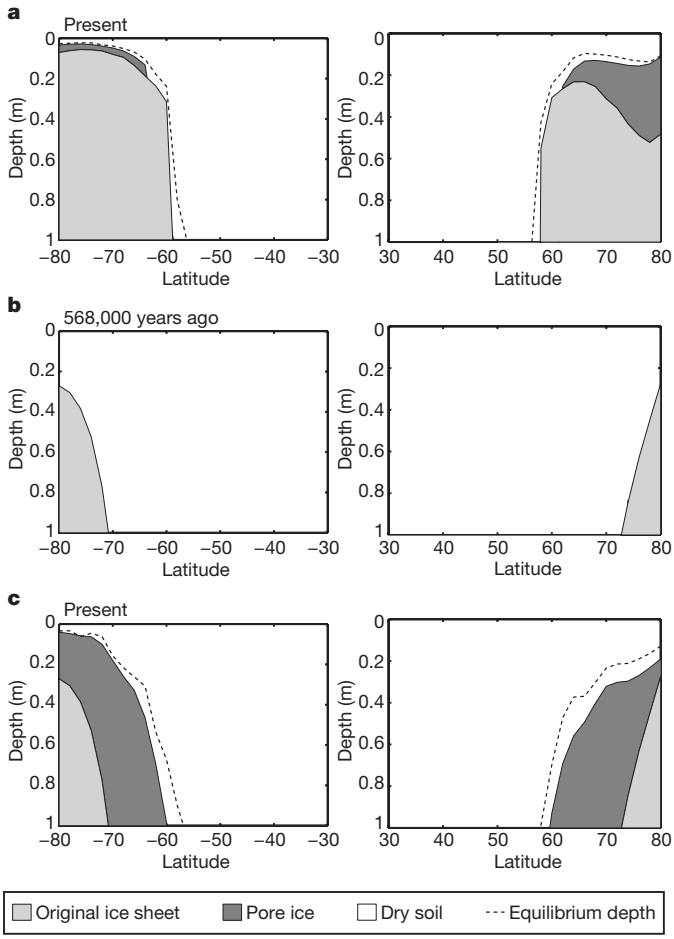


Figure 1 | Snapshots of the vertical ice distribution from model calculations. a, At present and with atmospheric humidity constant throughout history. **b**, For strongly varying atmospheric humidity, 568,000 years ago. **c**, For strongly varying atmospheric humidity, at present. Dashed lines show the present-day equilibrium depth.

[1]Institute for Astronomy and NASA Astrobiology Institute, 2680 Woodlawn Drive, University of Hawaii, Honolulu, Hawaii 96822, USA.

Figure 1a shows the resulting ice distribution, assuming constant atmospheric humidity, so that all ice movement is caused directly by temperature changes from orbital forcing. The partial pressure of water vapour on the surface is set to the present-day global average value of 0.13 Pa. There are three subsurface layers: (1) the original ice sheet remaining from the initial massive ice cover that formed by precipitation, (2) pore ice that forms by inward diffusion and constitutes at most 40% of the volume, and (3) dry soil. The depth to the ice closely follows the present-day equilibrium line, which attests to the rapidity of subsurface–atmosphere exchange. What is left from the original ice sheet closely matches the maximum equilibrium depth of the past five million years (not shown). Surprisingly, the latitudinal boundary is about as observed, ~60°. The ice has not retreated further poleward, because despite large obliquity changes and large changes in annual mean insolation[18], the annual mean temperature at this intermediate latitude changed little. Further poleward, temperature varied more over the last few million years, and the equilibrium ice table oscillated in depth. Trends of thermal conductivity of the soil with latitude account for the asymmetry between the southern and northern hemispheres.

Ice ages are not only driven directly by astronomical forcing via insolation changes, but indirectly through the supply of water from the polar cap. The amount of water that sublimes from the cap varies strongly with the planet's axis tilt, and as a result the atmospheric water vapour content is expected to vary greatly with time. The atmospheric vapour abundance may vary between zero, when the residual water cap is covered with carbon dioxide ice year round, and values estimated to be as high as 1,000 precipitable micrometres[1,2]. Figure 1b and c shows the resulting ice distribution at two instances in time. Figure 1b is for a recent obliquity minimum, the last time all pore ice had disappeared. In Fig. 1c, the original ice sheet has retreated further poleward, while the pore-ice layer created by diffusive back-filling extends close to equilibrium. A comparison of Fig. 1a with Fig. 1c reveals that martian ice ages are to a large extent only indirectly controlled by orbital forcing, through the effect of insolation on atmospheric vapour content.

Figure 2a shows the evolution of southern hemisphere subsurface ice with a strongly varying atmospheric vapour content as a function of time. After an extremely rapid retreat from the tropics, ice reforms in pore spaces at high obliquity (dark grey). The retreat of the original ice sheet (light grey) proceeds during dry low-obliquity periods and is independent of the model humidity at high obliquity. Ground ice forms and retreats with every major obliquity excursion; its maximum latitudinal extent depends sensitively on the maximum atmospheric humidity. The final stage of the evolution is that of Fig. 1c. Figure 2b shows the output and intake of water from the subsurface, summing latitudes 80° S to 80° N. The receding subsurface ice becomes a significant source of water vapour at low-obliquity periods that would otherwise be extremely dry. This could slow the retreat of the original ice sheet. At present, the subsurface contribution to the atmospheric humidity is nearly neutral, because the ice is close to equilibrium.

Existing inverse models of the elemental composition of the ground, from neutron and γ-ray spectroscopy, use a one- or two-layer vertical hydrogen distribution in the soil (in addition to the atmospheric layer), where each layer is assumed to have a homogeneous elemental composition. Figure 3 compares Gamma Ray Spectrometer Instrument Suite (GRS) measurements[19,20] (one-layer model) with the predicted amount of ice left over from the original ice sheet or reformed recently by inward diffusion. Although neither of the two mechanisms by itself could simultaneously account for the mass fraction and latitudinal boundary of the observed ice, their combination provides just enough ice at the right places. This climate scenario simultaneously explains the extent and density of ground ice, with a realistically large soil diffusivity. The constant humidity scenario, on the other hand, is inconsistent with the GRS measurements (Fig. 3). Two-layer GRS results for the burial depth and ice content[19,21] show that not only does the burial depth change with latitude, but so does the fraction of ice. Although the overall ice fraction is high, there is a latitude range where the ice content is small enough to be compatible with pore ice.

Two very different histories of atmospheric humidity have been considered: constant humidity (Fig. 1a) and strongly varying humidity from an exposed cap (Fig. 1b and c, and Fig. 2). These two histories bracket many other possibilities. The results depend on the history of atmospheric humidity (Fig. 3). Additional model calculations with different values for the diffusion coefficient and dust content give similar results, in part because the diffusive filling always closely follows the equilibrium depth.

The ice age scenario described here provides a comprehensive picture of how global scale glaciations may have evolved on Mars. It predicts that the top surface layer in the mid-latitudes has been

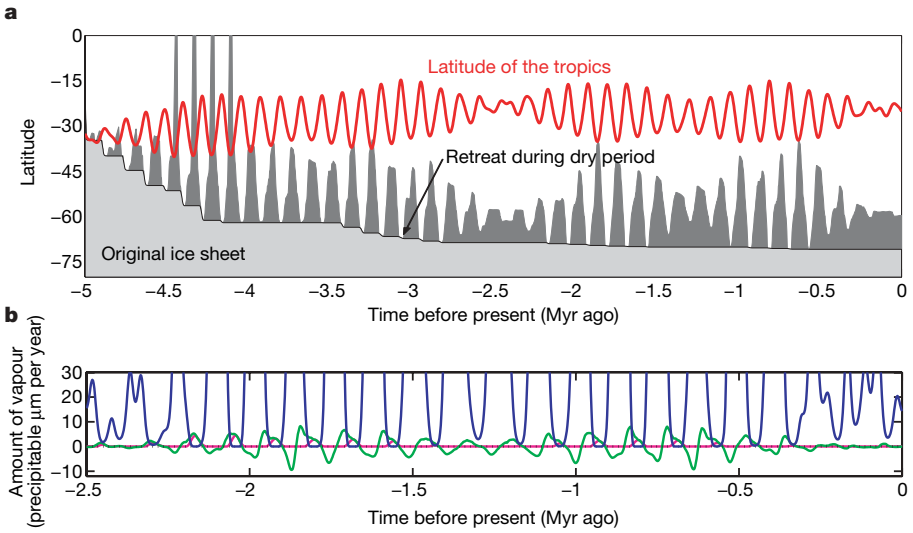


Figure 2 | Evolution of southern hemisphere ice over the past five million years for strongly varying atmospheric humidity. **a**, Latitudinal extent of ice in the topmost metre of soil. Shading as in Fig. 1. The red graph shows the latitude of the tropics, which equals (minus) the obliquity. Although only the obliquity is shown, the orbital eccentricity and longitude of perihelion also vary. **b**, The blue trace indicates the atmospheric humidity (originating from the north polar cap), the green trace indicates the output or intake of water by the subsurface, and the magenta trace indicates the contribution to the output from the retreat of the original ice sheet. Column-integrated amounts are given in units of thickness of a global ice layer.

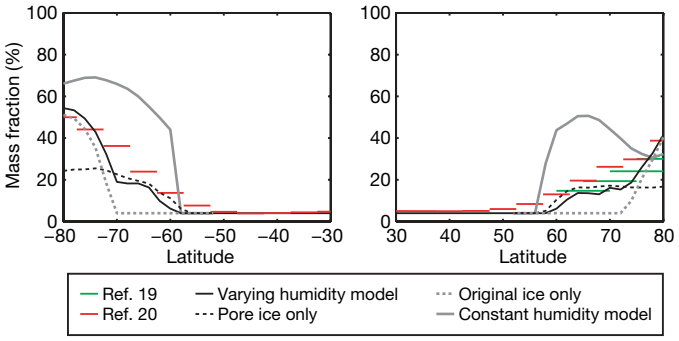


Figure 3 | Comparison with measurements by GRS onboard Mars Odyssey. Coloured bars show the ice (mass) fraction according to refs 19 and 20, in latitude bins, zonally averaged and assuming a homogeneous vertical distribution of hydrogen. The solid black and grey lines show the fraction of ice in the topmost metre of the soil according to the model described in the text and shown in Fig. 1a and c. The dry component is assumed to contain 4% water-equivalent hydrogen. The dashed and dotted lines pertain to the varying humidity scenario.

reworked numerous times during the past few million years[22,23], that much of the ice seen by GRS is diffusively formed pore-ice and less than half a million years old, but that an older ice sheet exists at higher latitudes. The pore-ice layers extend and retreat close to synchronously on both hemispheres. Over the last 2.5 million years the original ice sheets receded vertically by an estimated 60 cm (creating less than that in additional sublimation lag), considering only areas where it remains within the topmost metre today. Most of the ice loss, however, occurs at the retreating edge of the ice that lies below one metre, and the total ice volume change may be as high as 10^5 km^3, an amount that is still small compared to the volume of the permanent polar deposits. These movements are probably the counterpart of the layers observed in polar regions[24].

The dynamic nature of the ice sheets makes Mars an ideal system in which to test and expand our knowledge of astronomical climate forcing. A great deal could be learned about terrestrial ice ages from the study of martian ice stratigraphy—a longer, cleaner and simpler record than Earth's.

METHODS SUMMARY

The climate model is composed of models of surface and subsurface temperature, atmospheric humidity, and the retreat and growth of ground ice. The thermal model[11] solves the one-dimensional heat conduction equation with a semi-implicit numerical method and resolves annual and diurnal temperature cycles. The atmospheric humidity is controlled by the emission of water vapour from the north polar cap and is modelled according to Toon *et al.*[1]. The total atmospheric pressure is assumed to be constant.

Vapour transport calculations are intrinsically nonlinear, because of phase transitions, and hence require explicit numerical solvers for the diffusion equation, which are subject to stringent numerical stability requirements. Hence, explicit vapour diffusion models are orders of magnitude slower than a thermal model, a serious practical impediment for computations of ice evolution over long periods of Mars history. This limitation is overcome by employing net transport equations, as outlined in the following.

Previous work[11,16,25] has established a method to compute the retreat of ground ice efficiently. The vapour flux from the ice to the atmosphere is given by the difference in mean annual vapour density between the ice table and the surface, as in equation (1) of the Methods. A similar averaging method can be developed for the growth of pore ice. Detailed microphysical calculations have shown[11] that condensation of atmospherically derived ice in soil pores leads to accumulation of ground ice below a sharply defined interface. Below this interface, the diffusive flux is governed by the saturation vapour pressure, as shown in equation (4) in the Methods. The interface is located at a depth where the two inward fluxes (from the surface to the interface and from the interface downward) are equal, see equations (5) or (7). Instantaneous vapour density profiles are not required, and therein lies the computational efficiency.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 February; accepted 3 July 2007.

1. Toon, O. B., Pollack, J. B., Ward, W., Burns, J. A. & Bilski, K. The astronomical theory of climate change on Mars. *Icarus* **44,** 552–607 (1980).
2. Jakosky, B. M. & Carr, M. H. Possible precipitation of ice at low latitudes of Mars during periods of high obliquity. *Nature* **315,** 559–561 (1985).
3. Mellon, M. T. & Jakosky, B. M. The distribution and behavior of Martian ground ice during past and present epochs. *J. Geophys. Res.* **100,** 11781–11799 (1995).
4. Mischna, M. A., Richardson, M. I., Wilson, R. J. & McCleese, D. J. On the orbital forcing of martian water and CO_2 cycles: A general circulation model study with simplified volatile schemes. *J. Geophys. Res. E* **108,** 5062, doi:10.1029/2003JE002051 (2003).
5. Levrard, B., Forget, F., Montmessian, F. & Laskar, J. Recent ice-rich deposits formed at high latitudes on Mars by sublimation of unstable equatorial ice during low obliquity. *Nature* **431,** 1072–1075 (2004).
6. Forget, F., Haberle, R. M., Montmessin, F., Levrard, B. & Head, J. W. Formation of glaciers on Mars by atmospheric precipitation at high obliquity. *Science* **311,** 368–371 (2006).
7. Boynton, W. V. *et al.* Distribution of hydrogen in the near-surface of Mars: evidence for subsurface ice deposits. *Science* **297,** 81–85 (2002).
8. Feldman, W. C. *et al.* Global distribution of neutrons from Mars: results from Mars Odyssey. *Science* **297,** 75–78 (2002).
9. Mitrofanov, I. G. *et al.* Maps of subsurface hydrogen from the high-energy neutron detector, Mars Odyssey. *Science* **297,** 78–81 (2002).
10. Mellon, M. T., Feldman, W. C. & Prettyman, T. H. The presence and stability of ground ice in the southern hemisphere of Mars. *Icarus* **169,** 324–340 (2004).
11. Schorghofer, N. & Aharonson, O. Stability and exchange of subsurface ice on Mars. *J. Geophys. Res.* **110,** E05003, doi:10.1029/2004JE002350 (2005).
12. Mellon, M. T. & Jakosky, B. M. Geographic variations in the thermal and diffusive stability of ground ice on Mars. *J. Geophys. Res.* **98,** 3345–3364 (1993).
13. Hudson, T. L. *et al.* Water vapor diffusion in Mars subsurface environments. *J. Geophys. Res.* **112,** E05016, doi:10.1029/2006JE002815 (2007).
14. Leighton, R. B. & Murray, B. C. Behavior of carbon dioxide and other volatiles on Mars. *Science* **153,** 136–144 (1966).
15. Prettyman, T. H. *et al.* Composition and structure of the Martian surface at high southern latitudes from neutron spectroscopy. *J. Geophys. Res.* **109,** E05001, doi:10.1029/2003JE002139 (2004).
16. Mellon, M. T., Jakosky, B. M. & Postawko, S. E. The persistence of equatorial ground ice on Mars. *J. Geophys. Res.* **102,** 19357–19369 (1997).
17. Laskar, J. *et al.* Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* **170,** 343–364 (2004).
18. Ward, W. R. Climatic variations on Mars I. Astronomical theory of insolation. *J. Geophys. Res.* **79,** 3375–3386 (1974).
19. Litvak, M. L. *et al.* Comparison between polar regions of Mars from HEND/Odyssey data. *Icarus* **180,** 23–37 (2006).
20. Feldman, W. C. *et al.* The global distribution of near-surface hydrogen on Mars. *J. Geophys. Res.* **109,** E09006, doi:10.1029/2003JE002160 (2004).
21. Feldman, W. C. *et al.* Vertical distribution of hydrogen at high northern latitudes on Mars: The Mars Odyssey Neutron Spectrometer. *Geophys. Res. Lett.* **34,** L05201, doi:10.1029/2006GL028936 (2007).
22. Mustard, J. F., Cooper, C. D. & Rifkin, M. K. Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* **412,** 411–414 (2001).
23. Head, J. W., Mustard, J. F., Kreslavsky, M. A., Milliken, R. E. & Marchant, D. R. Recent ice ages on Mars. *Nature* **426,** 797–802 (2003).
24. Laskar, J., Levrard, B. & Mustard, J. F. Orbital forcing of the martian polar layered deposits. *Nature* **419,** 375–377 (2002).
25. Schorghofer, N. Theory of ground ice stability in sublimation environments. *Phys. Rev. E* **75,** 041201 (2007).

Acknowledgements I thank O. Aharonson and B. Jakosky for discussions and E. Pilger for computing help. This material is based upon work supported by the NASA Astrobiology Institute.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Correspondence and requests for materials should be addressed to the author (norbert@hawaii.edu).

METHODS

The climate model focuses on atmosphere–subsurface exchange of H_2O and it incorporates a new accelerated method for pore-ice calculations.

Thermal model. The thermal model is described in detail elsewhere[11]. It is run with a time step of 30 min in a layer 15 m thick. Temperatures are allowed to equilibrate over 14 Mars years and annual means are computed from the 15th Mars year. The model includes changes in thermal conductivity due to the presence of ground ice.

Fast numerical method for the retreat and growth of ground ice. Thermal models can take advantage of implicit or semi-implicit numerical methods that allow large time steps. Explicit vapour diffusion calculations are significantly slower than a thermal model, so here we exploit time-averaged equations for the retreat and growth of pore ice. The ice is assumed to recede vertically; internal deformations of the ice sheet are negligible, because its lateral extent is much larger than its depth.

Previous work[11,16,25] has established a method of computing the retreat of ground ice efficiently. The vapour flux from the ice to the atmosphere is given by:

$$J = -D(\langle\rho_{sv}(z_0)\rangle - \langle\rho(0)\rangle)/z_0 \quad (1)$$

where D is the diffusion coefficient of dry soil, ρ_{sv} the saturation vapour density, z_0 the depth of the ice table, and $\rho(0)$ the vapour density on the surface. Angle brackets indicate annual means. The averaging procedure not only provides significant computational advantages, but also eliminates dependences on several microphysical parameters, such as adsorption[11,25]. The rate of retreat of ground ice is given by $r = -J/\rho_{ice}$, where ρ_{ice} is the density of bulk ice, 927 kg m^{-3}. When a layer of dirty ice retreats, the speed at which the dry layer grows is:

$$\frac{dz_0}{dt} = \frac{1-\Phi_2}{1-\Phi_1} r \quad (2)$$

where Φ_1 is the porosity of the dry layer and Φ_2 the ice content of the ice layer. The geometric factor arises because dust in the ice is converted to porous sublimation till. Over a time step of $\Delta t = 250$ years, the dry layer grows from an initial thickness z_0 to a new thickness:

$$\sqrt{z_0^2 + 2\frac{1-\Phi_2}{1-\Phi_1} r z_0 \Delta t} \quad (3)$$

The retreat of pore ice can be handled similarly to the retreat of an ice layer.

A similar averaging method can be developed for the growth of pore ice. Detailed microphysical calculations have shown[11] that condensation of atmospherically derived ice in soil pores leads to accumulation of ground ice below a sharply defined interface. The depth z_0 of this interface can be determined as follows. Above the interface, the flux is given by equation (1). Below the interface, the diffusive flux is governed by the saturation vapour pressure:

$$J = -D(1-f)\langle\partial\rho_{sv}/\partial z\rangle \quad (4)$$

where D is the diffusion coefficient with all pores open, $f(z)$ is the volume fraction of pore space filled with ice, and z is the depth below the surface. Because f changes little within one Mars year, it can be taken outside the average. $(1-f)$ corresponds to constriction from partially ice-filled pores. The interface is located where the two inward fluxes (1) and (4) are equal:

$$(\langle\rho_{sv}(z_0)\rangle - \langle\rho(0)\rangle)/z_0 = (1-f)\langle\partial\rho_{sv}/\partial z\rangle|_{z_0} \quad (5)$$

With a Clausius–Clapeyron expression for the saturation vapour density, one obtains:

$$\frac{\partial\rho_{sv}}{\partial z} = \frac{\rho_{sv}}{T}\left(\frac{H}{RT} - 1\right)\frac{\partial T}{\partial z} \quad (6)$$

where T is the temperature, H is the enthalpy of sublimation, and R is the universal gas constant. With this expression, equation (5) becomes:

$$\frac{\langle\rho_{sv}(z_0)\rangle - \langle\rho(0)\rangle}{z_0} = (1-f)\left\langle\frac{\rho_{sv}}{T}\left(\frac{H}{RT} - 1\right)\frac{\partial T}{\partial z}\right\rangle\Bigg|_{z_0} \quad (7)$$

which is solved numerically for z_0. The instantaneous vapour density profile $\rho(z)$ is not needed, because ρ_{sv} is determined by temperature and $\rho(0)$ is prescribed by the surface humidity. When pore spaces are completely filled with ice ($f = 1$), equation (7) reduces to the equilibrium condition $\langle\rho_{sv}(z_0)\rangle = \langle\rho(0)\rangle$ and the flux vanishes. When the ice is stable but the pores are free of ice ($f = 0$), the interface is located at a depth below the equilibrium depth and the flux is inward. With time, the interface moves up to the equilibrium position, if the ice is stable; otherwise it retreats.

Atmospheric humidity. The atmospheric humidity is controlled by the emission of water vapour from the north polar cap and is modelled according to ref. 1. The model reproduces the present-day atmospheric vapour partial pressure on the surface of 0.13 Pa. For the past, the calculations assume an exposed cap of the same albedo and area as today's. On present-day Mars, vapour abundance generally decreases from the poles towards the equator, an effect which is neglected, because the frost point temperature depends only logarithmically on humidity; a model calculation with a gradient in partial pressure produced almost identical results. The total atmospheric pressure is assumed to be constant, which is supported by the notion that most CO_2 is stored in the atmosphere rather than in the polar caps; variations in CO_2 pressure would have a minor influence via changes in soil vapour diffusivity, atmospheric absorption and the size of the seasonal CO_2 cap.

LETTERS

The production of molecular positronium

D. B. Cassidy[1] & A. P. Mills Jr[1]

It has been known for many years that an electron and its antiparticle, the positron, may together form a metastable hydrogen-like atom, known as positronium or Ps (ref. 1). In 1946, Wheeler speculated[2] that two Ps atoms may combine to form the di-positronium molecule (Ps_2), with a binding energy[3] of 0.4 eV. More recently, this molecule has been studied theoretically[4]; however, because Ps has a short lifetime and it is difficult to obtain low-energy positrons in large numbers, Ps_2 has not previously been observed unambiguously[5]. Here we show that when intense positron bursts are implanted into a thin film of porous silica, Ps_2 is created on the internal pore surfaces. We found that molecule formation occurs much more efficiently than the competing process of spin exchange quenching, which appears to be suppressed in the confined pore geometry. This result experimentally confirms the existence of the Ps_2 molecule and paves the way for further multi-positronium work. Using similar techniques, but with a more intense positron source, we expect to increase the Ps density to the point where many thousands of atoms interact and can undergo a phase transition to form a Bose–Einstein condensate[6]. As a purely leptonic, macroscopic quantum matter–antimatter system this would be of interest in its own right, but it would also represent a milestone on the path to produce an annihilation gamma-ray laser[7].

The use of positron traps and accumulators[8] has recently made it possible to perform experiments with systems containing more than one positron[5]. In particular, interactions between Ps atoms have been studied by implanting intense positron pulses into a thin film of porous silica. This material was chosen because Ps atoms are formed quite efficiently in it (and similar substances) after positron irradiation[9]. Ps can be formed in a long-lived triplet state known as ortho-positronium (o-Ps decays with a 142 ns vacuum lifetime into three gamma-ray photons) or a short-lived singlet state known as para-positronium (p-Ps decays with a 125 ps vacuum lifetime into two photons)[1]. Ps is formed in the bulk material where it generally has a short lifetime (<1 ns), but atoms that diffuse into internal voids may become trapped therein and live for an appreciable fraction of the vacuum lifetime[10]. Under these circumstances the Ps decay rate is usually determined primarily by pick-off annihilation[11] following interactions with the pore walls. If the Ps density is high enough, o-Ps atoms may interact with each other, which can lead to one of two distinct processes: spin exchange quenching (SEQ, equation (1))[12] or the formation of Ps_2 (equation (2)):

$$\text{o-Ps} + \text{o-Ps} \leftrightarrow \text{p-Ps} + \text{p-Ps} + 2E_h \quad (1)$$

$$X + \text{o-Ps} + \text{o-Ps} \leftrightarrow X + Ps_2 + E_b \quad (2)$$

Here E_h is the energy difference between the o-Ps and p-Ps ground states (that is, the hyperfine energy, which is ~1 meV)[1] and X represents a third body. For SEQ to occur the energy $2E_h$ must be transferred either to the outgoing p-Ps atoms or to the surrounding medium. For Ps_2 formation to occur a third body is always required to conserve momentum. Both of these processes may result in the rapid annihilation of Ps and hence be detected via changes in Ps lifetime spectra. It is not, however, possible to distinguish between them using timing information alone.

The experimental arrangement was based on a Surko-type positron trap and accumulator[8] that has been described in detail elsewhere[13]. Positron pulses of about ten million particles with a sub-nanosecond time width were implanted into a 45% porous (porosity $P = 0.45$) silica film (made from tetraethoxysilane) that was 230 nm thick and had a 50-nm-thick non-porous capping layer[14]. The porous region contains interconnected pores with a diameter $d_{pore} \approx 4$ nm, and o-Ps that diffused into them had a lifetime of ~60 ns (ref. 10). With an incident beam areal density of 3×10^{10} cm^{-2} and a positronium fraction of ~10%, we estimate that the mean number of Ps atoms per pore was $\sim 10^{-5}$ (assuming a uniform distribution throughout the entire thickness of the porous region). However, the positronium atoms have a long diffusion length (~1 μm) and visit $\sim 10^4$ pores on average during their lifetime, so that the overall probability of two atoms interacting is around 10%.

All positrons implanted into the sample eventually annihilate with an electron. Immediately following implantation, many positrons annihilate with electrons directly, without forming positronium, producing a large pulse of gamma rays. Gamma rays resulting from the annihilation of p-Ps atoms are often indistinguishable from the direct annihilation[15]. Because the triplet Ps decays much more slowly, its presence is indicated by gamma rays that arrive some time after the incident positron pulse. The annihilation radiation was detected by a PbF_2 Cherenkov radiator optically coupled to a fast photomultiplier tube (PMT). Lifetime spectra were obtained by directly measuring the PMT anode voltage $V(t)$ with a fast oscilloscope[16].

Density-dependent changes in the Ps decay rate were investigated by recording lifetime spectra at five different beam areal densities n_{2D}. When the beam density is increased, the Ps lifetime is reduced owing to interactions between Ps atoms (a process we refer to as 'quenching'). The spectra were analysed to determine the delayed Ps fraction f_d, defined as:

$$f_d = \int_{20\text{ ns}}^{150\text{ ns}} V(t)\mathrm{d}t \Big/ \int_{-20\text{ ns}}^{150\text{ ns}} V(t)\mathrm{d}t \quad (3)$$

This parameter characterizes the amount of long-lived Ps present. For each group of five densities the mean value of f_d was calculated, and the deviation from that mean for each density, $\Delta f_d(n_{2D})$, was then used as a measure of the quenching. This procedure is equivalent to calculating the mean value of f_d at each density, but also takes into account the effects of small drifts that occur over the course of a run (typically ~12 h long). Figure 1 shows $\Delta f_d(n_{2D})$ for three representative temperatures; as in previous work, the expected linear density dependence of the quenching is observed[5].

The quenching signal would look the same whether it were due to Ps_2 formation or SEQ, because both of these mechanisms essentially convert long-lived triplet Ps into the short lived singlet state. We may, however, distinguish between the two mechanisms by considering the

[1]Department of Physics and Astronomy, University of California, Riverside, California 92521-0413, USA.

temperature dependence of the quenching. The basis for this differentiation is the fact that, neglecting three-body collisions, molecule formation is constrained to take place on a surface to conserve momentum, whereas SEQ is not. Ps may be thermally desorbed from surfaces, so we can control the fraction of atoms in the surface state via the temperature. This means that the temperature dependence of the quenching effect will be different for molecule formation and for SEQ; in the former case, heating will depopulate the surface states and therefore reduce the quenching, while in the latter case, heating would increase both the Ps density and the Ps–Ps interaction rate, which would increase the quenching. We observed that heating greatly reduces the quenching signal, unequivocally indicating molecule formation.

Figure 2a shows the temperature dependence of f_d. It is clear from the figure that the amount of long-lived Ps increases at higher temperatures, which we attribute to the thermal desorption of Ps from surface states. This leads to an increase in the amount of long-lived Ps present, because the lifetime of Ps on the surface is less than that in the voids. The data of Fig. 2a are divided into two distinct data sets; the difference between them is associated with the heating of the silica film to over 500 K for many hours (point 13). This was probably due to the thermal repair of positron trapping sites in the bulk material[17] or structural modifications caused by prolonged heating.

The solid lines in Fig. 2a are fits of an Arrhenius type, characteristic of thermally activated processes (see Methods). Such processes are typically parameterized by an activation energy E_a and a sticking (or accommodation) coefficient S, which is the probability that an atom remains on a surface immediately after impact. The thermal desorption of Ps from metallic surface states is well-known and has been extensively studied[18,19]. These activation energies are typically a few tenths of an electron volt and the sticking coefficients are close to unity, owing to the strong Coulomb interactions with the metallic electron gas[20]. There is also evidence to suggest that a Ps surface state exists on both crystalline[21,22] and amorphous SiO_2 (ref. 23) and our data are fully consistent with such a process. In this case S is very small because energy can only be exchanged with the surface by a weak coupling to phonons or other surface modes[24].

We quantify the quenching effect using the negative of the slope of $\Delta f_d(n_{2D})$, $Q \equiv -d\Delta f_d/dn_{2D}$. We used linear fits of $\Delta f_d(n_{2D})$, similar to those shown in Fig. 1, to all of the data to determine $Q(T)$, which is plotted in Fig. 2b. These data are separated into two distinct data sets in the same way as the data of Fig. 2a. The solid lines in Fig. 2b are fits similar to those of Fig. 2a (see Methods). Fitting the data in Fig. 2a yields activation energies (64 ± 23) and (83 ± 21) meV and sticking coefficients of $\log_{10}S = (-5.45 \pm 0.42)$ and (-5.34 ± 0.35) for the first and second data sets respectively. Similar fits were made to the data of Fig. 2b fixing the activation energy to be 74 meV, the average value obtained from the fits in Fig. 2a. The quality of the fits to all data sets reassures us that it is indeed appropriate to split the data into two sets, and that the model we use is essentially correct. The sticking coefficients obtained in this way for the first and second runs were $\log_{10}S = (-5.12 \pm 0.12)$ and (-4.88 ± 0.12). The small sticking coefficient ($S \approx 10^{-5}$) is consistent with a very light particle of mass $2m_e$ (where m_e is the mass of the electron) that can only lose energy via phonons on a surface made of SiO_2 molecules (mass $\approx 1.1 \times 10^5 m_e$).

We obtain the same sticking coefficient (within errors) from all four data sets, which means that the data are consistent with a single thermally activated process, namely the thermal desorption of Ps. Moreover, $f_d(T)$ and $Q(T)$ are essentially mirror images of one another, indicating that these quantities depend oppositely on the relative population of the Ps surface state. This indicates that the quenching process must be taking place on the pore surfaces, and we conclude that it is the formation of positronium molecules.

There are some other possible mechanisms that could, in principle, give rise to the signal we observe. To be consistent with the data, any such mechanism must involve two or more positrons, take place on the internal pore surfaces and result in a reduction of the amount of long-lived Ps.

We can immediately rule out the thermal dissociation of Ps_2 molecules as the source of $Q(T)$, because this would be unlikely to follow the same temperature dependence as the thermal desorption. Also, the activation energy we measure ($\sim$0.074 eV) is much less than the $\sim$0.4 eV Ps_2 binding energy[3]. Furthermore, half of the dissociated Ps atoms would be in a singlet state; these atoms would still contribute to the quenching signal and the maximum reduction in Q due to heating would therefore be 50%, which is not consistent with the data.

If the ''spur'' electrons[25] created in the silica by the incident positrons were able to interact with each other, it is possible that Ps formation could be inhibited in a manner that depends on the beam

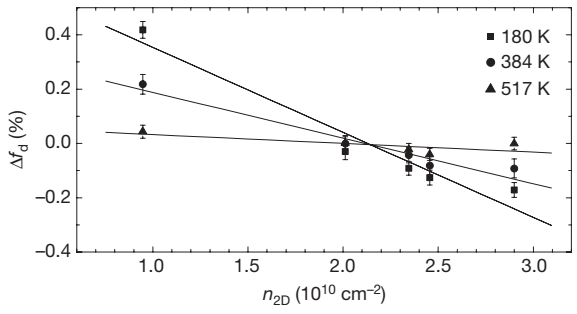


Figure 1 | Density dependence of the amount of long-lived Ps. The shift in f_d relative to the mean value for all beam densities is shown for three representative temperatures. The solid lines are linear fits used to determine the parameter Q as described in the text. The 1σ error bars are determined by the distribution of sets of at least 50 individual measurements each.

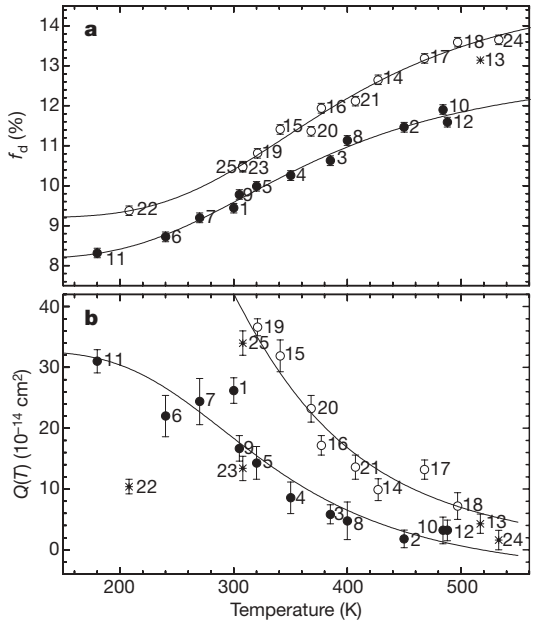


Figure 2 | Temperature dependence of f_d and Q. **a**, f_d as a function of temperature measured using a beam density of 0.9×10^{10} cm^{-2}. The 1σ error bars were determined by requiring the χ^2 value per degree of freedom for the fits to be close to unity. **b**, Q as a function of temperature obtained by fitting the density dependence of f_d. The 1σ error bars are from the fits used to obtain Q as in Fig. 1. The data points are labelled in the order in which they were taken. The lines are fits of equations (4) to the data. The filled and open circles are data from the first and second runs respectively, whereas the star symbols represent data ignored in the fits. Point 13 was apparently in transition between the low- and high-f_d runs. Points 22–25 were obtained using an improperly adjusted positron accumulator that yielded on average about twice as many positrons per pulse as the rest of the data. For these runs the beam intensity was unstable and actually of lower (and uncertain) areal density. These data yield reliable values for f_d, but we do not trust them to give usable measures of Q.

density. However, the spur electron cloud radius is around 10 nm, so this is extremely unlikely at the present beam densities, especially as only surface effects would be consistent with our observations.

If the Ps density were greatly enhanced on the surface relative to that in the pores, then the thermal desorption of Ps could lead to a reduction in Q even in the absence of Ps_2 molecule formation. For pore sizes of the order of the Ps thermal de Broglie wavelength we may describe the surface state Ps as the ground state of the cavity and the free Ps as the first excited state. We can then estimate the change in the mean Ps density that is due to desorption from the surface. For a reasonable approximation of the surface potential (4 atomic units wide, 1.8 eV deep[22]) and for a spherical cavity 4 nm in diameter we find that the surface density is actually less than the free Ps density by a factor of two. SEQ on the surface therefore cannot explain the temperature dependence of Q. Moreover, for a surface interaction where momentum is conserved, the Ps_2 formation rate compared to that for SEQ for a given density is highly favoured because the latter proceeds through a virtual Ps_2 state[26]. Thus, while there could be some SEQ between Ps atoms in the surface state, it is a second-order process and molecule formation is far more likely.

It is surprising that there is so little SEQ at higher temperatures. If we assume that there is no Ps_2 formation at all above 500 K, then the fits to the data of Fig. 2b imply that at room temperature Ps_2 formation is roughly ten times more likely than SEQ. This is consistent with SEQ being a second-order process, but one might expect that at higher temperatures when molecule formation is suppressed and the Ps density is increased SEQ would occur fairly efficiently. That it does not implies that spin exchanging collisions in the voids are suppressed.

Spin exchange from triplet to singlet Ps requires the transfer of the 1 meV triplet hyperfine energy to the kinetic energies of the final-state singlet Ps atoms (see equation (1)). Because the lowest energy surface and volume states are discrete (with energies of about 35 meV in a 4-nm-diameter pore), there are no nearby localized Ps states available to accommodate the extra energy. Thus, spin exchange will be highly suppressed not only in the volumes of a confined geometry, but also on the surfaces because it can only occur if accompanied by a phonon to conserve energy or by an improbable coupling to a chance pair of delocalized Ps states of the correct energy. This idea is supported by measurements made on a different sample[5] in which the pores were aligned to form long one-dimensional tubes; this type of pore geometry allows a continuous distribution of momentum in one dimension and would not be expected to suppress SEQ. With this sample we observed no thermal desorption of Ps, but a much stronger quenching signal, $Q = (154 \pm 5) \times 10^{-14}\,\mathrm{cm}^2$, that had no apparent temperature dependence. Further studies using samples of various geometries would help to elucidate the details of this process, but it is clear that in the disordered geometry of the porous silica film the observed quenching signal cannot be explained by SEQ.

The data presented here constitute evidence that we have observed molecular positronium. Our experiments suggest that a single-crystal quartz surface might be a good source of Ps_2 molecules, and work is underway to make a direct observation using a laser to excite a resonant molecular transition[27,28].

METHODS SUMMARY

We modelled the Ps thermal desorption assuming that Ps atoms are localized on the surfaces of single pores and that they can occupy one of M equivalent sites per unit area L^2. The yield as a function of temperature $Y(T)$ of Ps atoms annihilating in the voids depends on the ratio of the desorption rate to the total annihilation rate of the Ps on the surface. Following ref. 19 we write:

$$Y(T) = \left[1 + (4/e)(\lambda_{\mathrm{Ps}}^2 M/L^2) S^{-1} (\gamma \lambda_{\mathrm{Ps}}/\bar{\nu}) \exp(E_{\mathrm{a}}/kT)\right]^{-1} \quad (4)$$

where γ is the annihilation rate of Ps on the surface, the thermal de Broglie wavelength of a positronium atom is $\lambda_{\mathrm{Ps}} = \sqrt{\pi\hbar^2/m_{\mathrm{e}}kT}$, the thermal velocity $\bar{\nu} = \sqrt{4kT/\pi m_{\mathrm{e}}}$, S is the Ps sticking coefficient[29] and E_{a} is the activation energy for thermal desorption of the bound Ps. We assume $M/L^2 = (\pi d_{\mathrm{pore}}^2)^{-1} \approx 2 \times 10^{12}\,\mathrm{cm}^{-2}$ and the surface state lifetime may be estimated from the ~3.5% overlap of the Ps wavefunction with the SiO_2 surface[22]. Assuming the annihilation rate of Ps in the bulk material is $3\,\mathrm{ns}^{-1}$, the surface state decay rate $\gamma \approx 10^8\,\mathrm{s}^{-1}$. Then we obtain:

$$Y(T) = \left[1 + (300\mathrm{K}/T)^2 S^{-1} (7.2 \times 10^{-7}) \exp(E_{\mathrm{a}}/kT)\right]^{-1} \quad (5)$$

The solid lines in Fig. 2a and b are fits using the functions:

$$f_{\mathrm{d}}(T) = f_0 + f_1 Y(T) \quad \text{and} \quad Q(T) = Q_0 + Q_1[1 - Y(T)] \quad (6)$$

respectively. The subscripts 0 and 1 refer to the non-thermal and thermal components (respectively) of f_{d} and Q.

Received 11 June; accepted 17 July 2007.

1. Charlton, M. & Humberston, J. W. *Positron Physics* (Cambridge Univ. Press, Cambridge, 2001).
2. Wheeler, J. A. Polyelectrons. *Ann. NY Acad. Sci.* **48,** 219–238 (1946).
3. Hylleraas, E. A. & Ore, A. Binding energy of the positronium molecule. *Phys. Rev.* **71,** 493–496 (1947).
4. Schrader, D. M. Symmetry of dipositronium Ps_2. *Phys. Rev. Lett.* **92,** 043401 (2004).
5. Cassidy, D. B. *et al.* Experiments with a high-density positronium gas. *Phys. Rev. Lett.* **95,** 195006 (2005).
6. Platzman, P. M. & Mills, A. P. Jr. Possibilities for Bose condensation of positronium. *Phys. Rev. B* **49,** 454–458 (1994).
7. Mills, A. P. Jr, Cassidy, D. B. & Greaves, R. G. Prospects for making a Bose-Einstein condensed positronium annihilation gamma ray laser. *Mater. Sci. Forum* **445,** 424–429 (2004).
8. Greaves, R. G. & Surko, C. M. Emerging physics and technology of antimatter plasmas and trap-based beams. *Phys. Plasmas* **11,** 2333–2348 (2004).
9. Paulin, R. & Ambrosino, G. Annihilation libre de l'ortho-positronium formé dans certaines poudres de grande surface specifique. *J. Phys. (Paris)* **29,** 263–270 (1968).
10. Gidley, D. W. *et al.* Positronium annihilation in mesoporous thin films. *Phys. Rev. B* **60,** R5157–R5160 (1999).
11. Brandt, W., Berko, S. & Walker, W. W. Positronium decay in molecular substances. *Phys. Rev.* **120,** 1289–1295 (1960).
12. Deutsch, M. Evidence for the formation of positronium in gases. *Phys. Rev.* **82,** 455–456 (1951).
13. Cassidy, D. B., Deng, S. H. M., Greaves, R. G. & Mills, A. P. Jr. Accumulator for the production of intense positron pulses. *Rev. Sci. Instrum.* **77,** 073106 (2006).
14. Tanaka, H. K. M., Kurihara, T. & Mills, A. P. Jr. Evaluation of the diffusion barrier continuity on porous low-k films using positronium time of flight spectroscopy. *Phys. Rev. B* **72,** 193408 (2005).
15. Saito, H. & Hyodo, T. Direct measurement of the parapositronium lifetime in α-SiO_2. *Phys. Rev. Lett.* **90,** 193401 (2003).
16. Cassidy, D. B., Deng, S. H. M., Tanaka, H. K. M. & Mills, A. P. Jr. Single shot positron annihilation lifetime spectroscopy. *Appl. Phys. Lett.* **88,** 194105 (2006).
17. Cassidy, D. B. & Mills, A. P. Jr. Radiation damage in a-SiO_2 exposed to intense positron pulses. *Nucl. Instrum. Methods B* **262,** 59–64 (2007).
18. Mills, A. P. Jr. Thermal activation measurement of positron binding energies at surfaces. *Solid State Commun.* **31,** 623–626 (1979).
19. Chu, S., Mills, A. P. Jr & Murray, C. A. Thermodynamics of positronium thermal desorption from surfaces. *Phys. Rev. B* **23,** 2060–2064 (1981).
20. Martin, Th, Bruinsma, R. & Platzman, P. M. Adsorption of positronium on metal surfaces: theory. *Phys. Rev. B* **43,** 6466–6473 (1991).
21. Sferlazzo, P., Berko, S. & Canter, K. F. Experimental support for physisorbed positronium at the surface of quartz. *Phys. Rev. B* **32,** 6067–6070 (1985).
22. Saniz, R., Barbiellini, B., Platzman, P. M. & Freeman, A. J. Physisorption of positronium on quartz surfaces. Preprint at ⟨http://arxiv.org/abs/0705.0037⟩ (2007).
23. Kim, S. M. & Buyers, W. J. L. Positronium-surface interaction in the pores of the vycor glass. *J. Phys. C* **11,** 101–109 (1978).
24. He, C. *et al.* Evidence for pore surface dependent positronium thermalization in mesoporous silica/hybrid silica films. *Phys. Rev. B* **75,** 195404 (2007).
25. Mogensen, O. E. Spur reaction model of positronium formation. *J. Chem. Phys.* **60,** 998–1004 (1974).
26. Mills, A. P. Jr. in *Positron Spectroscopy of Solids* (eds Dupasquier, A. & Mills, A. P. Jr) 209–258 (IOS Press, Amsterdam, 1995).
27. Mills, A. P. Jr. Chemistry and physics with many positrons. *Rad. Phys. Chem.* **76,** 76–83 (2007).
28. Varga, K., Usukura, J. & Suzuki, Y. Second bound state of the positronium molecule and biexcitons. *Phys. Rev. Lett.* **80,** 1876–1879 (1998).
29. Langmuir, I. The adsorption of gases on plane surfaces of glass, mica and platinum. *J. Am. Chem. Soc.* **40,** 1361–1403 (1918).

Acknowledgements We gratefully acknowledge R. G. Greaves for discussions and H. K. M. Tanaka for providing the porous silica film. This work was supported in part by the National Science Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.B.C. (cassidy@physics.ucr.edu).

LETTERS

Coupling of surface temperatures and atmospheric CO_2 concentrations during the Palaeozoic era

Rosemarie E. Came[1], John M. Eiler[1], Ján Veizer[2], Karem Azmy[3], Uwe Brand[4] & Christopher R. Weidman[5]

Atmospheric carbon dioxide concentrations seem to have been several times modern levels during much of the Palaeozoic era (543–248 million years ago), but decreased during the Carboniferous period to concentrations similar to that of today[1–3]. Given that carbon dioxide is a greenhouse gas, it has been proposed that surface temperatures were significantly higher during the earlier portions of the Palaeozoic era[1]. A reconstruction of tropical sea surface temperatures based on the $\delta^{18}O$ of carbonate fossils indicates, however, that the magnitude of temperature variability throughout this period was small[4], suggesting that global climate may be independent of variations in atmospheric carbon dioxide concentration. Here we present estimates of sea surface temperatures that were obtained from fossil brachiopod and mollusc shells using the 'carbonate clumped isotope' method[5]—an approach that, unlike the $\delta^{18}O$ method, does not require independent estimates of the isotopic composition of the Palaeozoic ocean. Our results indicate that tropical sea surface temperatures were significantly higher than today during the Early Silurian period (443–423 Myr ago), when carbon dioxide concentrations are thought to have been relatively high, and were broadly similar to today during the Late Carboniferous period (314–300 Myr ago), when carbon dioxide concentrations are thought to have been similar to the present-day value. Our results are consistent with the proposal that increased atmospheric carbon dioxide concentrations drive or amplify increased global temperatures[1,6].

The link between atmospheric CO_2 concentrations and Earth surface temperatures is central to our understanding of environmental change at many times in Earth history[7]. Among the most puzzling times for our understanding of the climatic consequences of CO_2 is the Palaeozoic era (the period between 543 and 248 Myr ago) that saw the emergence and diversification of the major classes of large-bodied animal and plant life forms.

Modelled atmospheric CO_2 concentrations between the mid-Cambrian and latest Silurian (530–417 Myr ago; that is, the early Palaeozoic) were 12–17 times higher than the modern atmosphere, and were followed by far lower levels (comparable to the modern atmosphere) during the Carboniferous period (354–290 Myr ago; that is, the late Palaeozoic)[1,2]. These model estimates are supported by independent geochemical proxies of atmospheric CO_2 (refs 8–10). In addition, the large decrease in atmospheric CO_2 that these models propose for the beginning of the Carboniferous period is a plausible cause of the extensive Carboniferous glaciation[11–13].

However, northern Africa experienced extensive glaciation during the Late Ordovician and Early Silurian[12,13], when atmospheric CO_2 is inferred to have been 12–17 times modern values[1,2,10]. This suggests that either model reconstructions of atmospheric CO_2 levels are prone to large errors or that climate can vary dramatically, independent of variations in atmospheric CO_2. Furthermore, Veizer *et al.*[4] have reconstructed tropical shallow-marine temperatures during the Palaeozoic by applying the oxygen isotope carbonate-water thermometer to well-preserved carbonate fossils from sediments deposited in shallow water at low latitude (less than about 30° of latitude). Their results suggest that tropical shallow-marine temperatures were similar to each other and within ~5 °C of modern conditions during times of high inferred atmospheric CO_2, such as the Silurian, and during times of lower inferred atmospheric CO_2, such as the Carboniferous. The Phanerozoic temperature trend implied by Veizer *et al.*[4] has four 'icehouse/greenhouse' modes and resembles the sedimentological and palaeontological climate reconstruction of Scotese (see ref. 14 and www.scotese.com/climate.htm). Veizer *et al.*[4] suggest on this basis that global climate is not well coupled with atmospheric CO_2 concentrations over the timescale of the Phanerozoic eon[4].

This debate regarding climatic conditions during the Palaeozoic suffers from two uncertainties. First, geologic evidence for the spatial and temporal distribution of sediments and fossils provides qualitative constraints on climate, but cannot be easily translated into a measure of global temperature and therefore does not clearly and directly test models of global climate. Second, oxygen isotope constraints on surface temperature are vulnerable to artefacts from diagenetic or burial-metamorphic overprints[4,15] and require assumptions or independent constraints on the oxygen isotopic compositions of the waters in which carbonate fossils grew. These issues have led to decades-long uncertainty as to whether the systematic temporal variations in oxygen isotope compositions of Phanerozoic marine carbonate fossils reflects climate change, variation in the $\delta^{18}O$ of sea water, or post-depositional alteration[15,16].

We address these issues by applying the carbonate clumped-isotope thermometer to aragonite and low-Mg calcite fossils of Palaeozoic age. This thermometer examines ordering, or 'clumping', of ^{13}C and ^{18}O into bonds with each other in the carbonate mineral lattice. This isotope effect is temperature dependent, and can be examined by analysis of $^{13}C^{18}O^{16}O$ in CO_2 released from carbonates by phosphoric acid digestion. Importantly, it provides a temperature constraint that depends only on the isotopic composition of carbonate and is independent of the isotopic composition of the water in which the carbonate grew[5] (see Methods). Furthermore, our approach permits us to estimate the $\delta^{18}O$ of sea water on the basis of known growth temperatures and $\delta^{18}O$ values of carbonate fossils.

We examined two suites of relatively well preserved carbonate shells of shallow-water marine organisms that lived at palaeolatitudes within 20° of the Equator: (1) early Silurian brachiopods consisting of low-Mg calcite, collected from the Telychian-age Jupiter Formation on Anticosti Island, Canada[17,18]; and (2) Carboniferous (Middle Pennsylvanian) aragonitic molluscs, collected from the

[1]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA. [2]Ottawa-Carleton Geoscience Centre, University of Ottawa, Ottawa, Ontario KIN 6N5, Canada. [3]Department of Earth Sciences, Memorial University of Newfoundland, St John's, Newfoundland A1B 3X5, Canada. [4]Department of Earth Sciences, Brock University, St Catharines, Ontario L2S 3A1, Canada. [5]Waquoit Bay National Estuarine Research Reserve, Waquoit, Massachusetts 02536, USA.

Boggy Formation, in southern Oklahoma, USA[19]. Both suites include equal numbers of samples that appear, on independent evidence (visual, microscopic[18], X-ray diffraction and/or trace-element analysis[20,21]), to be well preserved, and samples that appear to be moderately altered by post-depositional processes. These two sub-sets of each suite were selected so that we could systematically examine the effects of alteration on the isotopic record (see Supplementary Information for further details).

Pennsylvanian samples exhibit a positive correlation between $\delta^{13}C$ and $\delta^{18}O$ values (Table 1 and Fig. 1). The high-$\delta^{18}O$, high-$\delta^{13}C$ end of this trend is associated with elevated Fe abundances, Mn abundances and/or proportions of secondary calcite, suggesting that post-depositional alteration caused increases in $\delta^{13}C$ and $\delta^{18}O$ in this suite. The direction of this trend is contrary to common expectations that alteration leads to decreases in $\delta^{18}O$ and $\delta^{13}C$ (refs 15, 16, 21), but the mineralogical, textural and trace element attributes of the studied samples argue for such an interpretation. Silurian samples exhibit little variation in $\delta^{18}O$ but significant variability in $\delta^{13}C$ (Table 1 and Fig. 1). On the basis of visual evidence for recrystallization[18], lower $\delta^{13}C$ values are associated with increasing post-depositional alteration (this result, though supported by relatively straightforward observations, is also contrary to common inferences regarding the isotopic effects of diagenesis and burial metamorphism).

The apparent temperatures of carbonate growth based on clumped isotope thermometry and the calculated $\delta^{18}O$ values of water in equilibrium with our samples at those apparent temperatures are presented in Table 1 and Fig. 1. Pennsylvanian samples exhibit a positive correlation between temperature and $\delta^{18}O$ of water, and a clear association of altered samples with higher temperatures and higher water $\delta^{18}O$. The data reinforce our interpretation of the correlation between $\delta^{18}O$ and $\delta^{13}C$ for these samples, and indicate that the low-Fe, low-Mn, aragonite-rich sub-set of this suite (samples B81-21, B81-18 and B81-06x) most closely preserve their depositional isotopic compositions. Silurian samples yield a bimodal distribution of apparent temperatures and values of water $\delta^{18}O$: all but one of the nominally unaltered samples (based on visual inspection; that is, samples A1380b-30, A1391b-31, A1391b-06 and A1356a-37) group tightly around a mean temperature near $\sim$35 °C. Visibly altered samples exhibit a positive correlation between temperature and $\delta^{18}O$ of water, and an association of altered samples with higher temperatures and higher water $\delta^{18}O$. These data also reinforce our interpretation that alteration is associated with low carbonate $\delta^{13}C$ in this sample suite.

We suggest that the average apparent temperatures and water $\delta^{18}O$ values recorded by the least altered subset of each sample suite best represent the original fossil growth conditions. These data yield nominal growth conditions of 24.9 ± 1.7 °C and seawater $\delta^{18}O$ of -1.6 ± 0.5‰ (both ± 1 s.e.) for the Middle Pennsylvanian, and 34.9 ± 0.4 °C and seawater $\delta^{18}O$ of -1.2 ± 0.1‰ for the Early Silurian. It is possible that even these averages are influenced by subtle post-depositional alteration and thus over-estimate depositional temperatures and seawater $\delta^{18}O$ values. It is not clear how one could ever strictly disprove this possibility. Nevertheless, all samples that contribute to these averages pass conventional criteria for a high level of preservation, each group is homogeneous within analytical precision, the Silurian data are consistent with previous measurements of the $\delta^{18}O$ values of well-preserved conodont fossils[22], and both of our suites yield seawater $\delta^{18}O$ similar to that of the modern ocean, after adding back the water currently stored as polar ice. This is in accord with models of the global water isotopic budget that suggest nearly constant $\delta^{18}O$ of sea water throughout the Phanerozoic eon[23,24] (but see ref. 25 for an alternative model that permits variations in the $\delta^{18}O$ of sea water). Finally, we also note that average growth temperatures and water $\delta^{18}O$ values for nominally unaltered samples from each suite may be influenced by ecological variability, and thus may be offset from true averages for their respective palaeolatitudes and ages. This possibility could be investigated further through more detailed studies covering a wider range of locations and depositional conditions for a given time period.

Our results provide a basis for discriminating between previous competing hypotheses regarding the character of Palaeozoic climate change and the $\delta^{18}O$ of the Phanerozoic ocean. First, we find that when atmospheric CO_2 is inferred to have been highly elevated compared to modern levels—that is, during the Early Silurian—shallow-marine temperatures were markedly elevated, and when atmospheric CO_2 was nearly as low as modern values—during the Middle Pennsylvanian—shallow-marine temperatures were similar to modern values[1,2]. This result is consistent with the proposition that variations in atmospheric CO_2 concentration from the Silurian to the Pennsylvanian drove large variations in Earth surface temperatures[1] (Fig. 2; but note this raises a new question as to how such warm temperatures could be consistent with geological evidence for high-latitude glaciation during the earliest Silurian[12,13]). More generally, our results are consistent with the hypothesis that elevated CO_2 concentrations are capable of producing Earth surface temperatures substantially (5–11 °C) higher than modern values. Second, our results support previous arguments that the $\delta^{18}O$ of sea water has varied within a narrow range throughout the Phanerozoic eon[23,24] and argue against suggestions that it was several per mil lower during the Palaeozoic[15,17]. Although there are many differences between Palaeozoic and modern climates, the suggestion our results give of a link between increased CO_2 and a large temperature increase provides a point of reference for models of projected climate change associated with currently rising concentrations of atmospheric greenhouse gases.

Table 1 | Stable isotope and temperature data

Sample ID	Species	$\delta^{18}O_c$ (‰; PDB)	$\delta^{13}C_c$ (‰; PDB)	Δ_{47} (‰)	T (°C)	$\delta^{18}O$ of water (‰; SMOW)
Unaltered Middle Pennsylvanian samples						
B81-21	*Domatoceras* sp.	−3.15	0.49	0.66	21.6	−2.77
B81-21	*Domatoceras* sp.	−3.26	0.55	0.65	24.0	−2.37
B81-21	*Domatoceras* sp.	−2.81	0.52	0.67	19.0	−2.97
B81-18	*Metacoceras cornutum*	−2.38	1.92	0.64	26.8	−0.92
B81-18	*Metacoceras cornutum*	−2.33	1.97	0.64	26.1	−0.99
B81-06x	*Orthoceras unicamera*	−2.99	2.00	0.65	25.1	−1.86
B81-06x	*Orthoceras unicamera*	−2.74	2.02	0.62	30.2	−0.58
B81-06x	*Orthoceras unicamera*	−2.41	2.02	0.64	27.0	−0.89
B81-06x	*Orthoceras unicamera*	−2.67	2.03	0.65	24.6	−1.65
Middle Pennsylvanian samples suspected of alteration						
B0-01	Cephalopod fragment	−1.43	3.25	0.64	26.8	0.04
B0-01	Cephalopod fragment	−1.43	3.22	0.63	27.9	0.25
B0-04	Cephalopod fragment	−1.13	3.50	0.65	23.4	−0.37
B0-04	Cephalopod fragment	−1.07	3.52	0.65	24.2	−0.14
Altered Middle Pennsylvanian samples						
B81-27	Cephalopod fragment	−0.25	5.60	0.63	29.3	1.72
B81-27	Cephalopod fragment	−0.28	5.61	0.63	29.2	1.69
B81-09	*Domatoceras* sp.	−0.76	3.74	0.62	30.5	1.46
B81-09	*Domatoceras* sp.	−0.78	3.66	0.63	29.7	1.27
B81-10	*Domatoceras* sp.	−0.28	3.52	0.61	34.5	2.74
B81-10	*Domatoceras* sp.	−0.73	3.51	0.62	32.0	1.79
B81-40	*Pseudorthoceras knoxense*	0.28	6.36	0.60	37.0	3.79
Unaltered Silurian samples						
A1380b-30	*Pentamerus* sp.	−5.34	1.94	0.61	33.7	−1.50
A1391b-31	*Pentamerus* sp.	−5.00	2.08	0.60	34.8	−0.95
A1391b-31	*Pentamerus* sp.	−4.94	2.03	0.60	35.3	−0.79
A1391b-06	*Pentamerus* sp.	−5.11	1.78	0.60	35.7	−0.88
A1391b-06	*Pentamerus* sp.	−5.12	1.71	0.61	34.2	−1.19
A1356a-37	*Pentamerus* sp.	−5.63	1.82	0.60	35.8	−1.39
Silurian samples suspected of alteration						
A1356a-07	*Pentamerus* sp.	−5.07	1.73	0.57	42.9	0.48
A1356a-07	*Pentamerus* sp.	−5.10	1.71	0.57	43.9	0.62
Altered Silurian samples						
A-551 alt	*Stricklandia planirostrata*	−5.22	−1.13	0.50	63.8	3.89
A-958	*Ehlersella davidsoni*	−5.12	0.18	0.51	62.4	3.75
A-958-2 alt	*Ehlersella davidsoni*	−5.20	−0.03	0.51	61.3	3.49
A-958-3	*Ehlersella davidsoni*	−4.88	0.86	0.54	53.0	2.43
A-958-4 alt	*Ehlersella davidsoni*	−4.88	0.21	0.52	56.7	3.07

$\delta^{18}O_c$ and $\delta^{13}C_c$ are the $\delta^{18}O$ and $\delta^{13}C$ of carbonate. The Δ_{47} is defined as the difference in ‰ between the measured 47/44 ratio of the sample and the 47/44 ratio expected for that sample if its stable carbon and oxygen isotopes were randomly distributed among all isotopologues of CO_2 (ref. 27) (see Methods) Here T is the Δ_{47}-derived temperature in degrees Celsius.

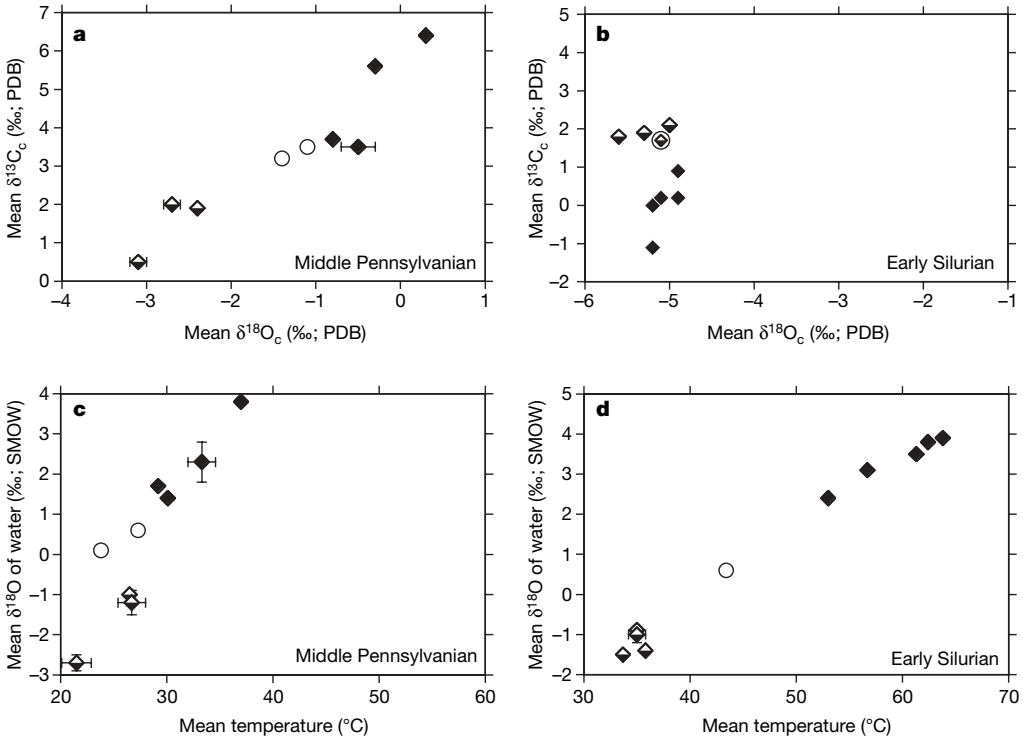


Figure 1 | Isotopic compositions and inferred crystallization temperatures of Pennsylvanian and Silurian fossils. **a**, Mean $\delta^{18}O$ of carbonate versus mean $\delta^{13}C$ of carbonate (PDB) for Pennsylvanian aragonitic molluscs; **b**, same for Silurian calcitic brachiopods. **c**, Mean Δ_{47}-derived temperatures versus the calculated $\delta^{18}O$ (SMOW) of sea water and/or diagenetic waters for Pennsylvanian aragonitic molluscs; **d**, same for Silurian calcitic brachiopods. In all panels, well-preserved samples are represented by half-filled diamonds; diagenetically altered samples by filled diamonds; and samples suspected of alteration by open circles. Error bars represent ±1 s.e. on replicate analyses.

Note that the discrepancy between our results and the more subtle temperature variations in the reconstruction of Veizer *et al.* (Fig. 2) primarily reflects the fact that this previous study assumed large ice volume variations in the $\delta^{18}O$ of the ocean, whereas our data suggest such changes were minimal. Thus, our results suggest both positive and negative things about previous attempts at palaeothermometry based on the $\delta^{18}O$ of Palaeozoic carbonate fossils. On the one hand, we confirm that carefully selected fossils of this age are characterized by isotopic compositions that reflect their conditions of deposition (that is, the previously proposed criteria for identifying well-preserved samples are mostly predictive for samples that have experienced minimal re-crystallization and/or isotopic exchange). On the other hand, these earlier studies suggested that the variable and, on average, low values of $\delta^{18}O$ of well-preserved Palaeozoic fossils should be interpreted as evidence for secular variation in the $\delta^{18}O$ of sea water in an ocean that varied little in average temperature, whereas our data suggest the converse (Figs 2 and 3); that is, the results of this study support the data but contradict the interpretation of Veizer *et al.*

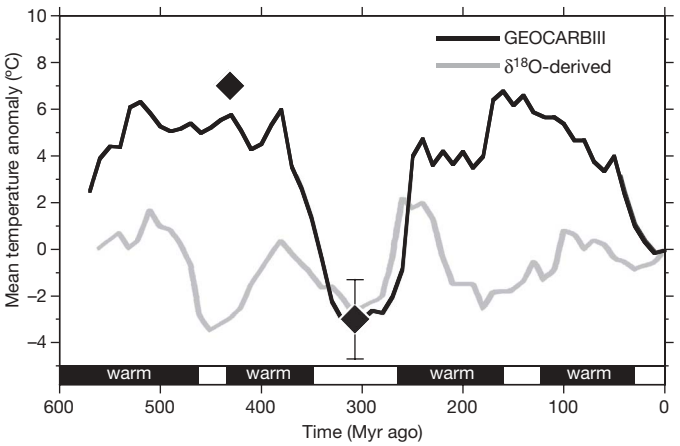


Figure 2 | Estimates of tropical temperature anomalies relative to today. The grey[4] curve represents temperature estimates based on $\delta^{18}O$ of well-preserved carbonate fossils from palaeotropical seas; the black[2] curve represents GEOCARBIII model estimates of mean global temperatures based on reconstructions of atmospheric CO_2 levels; black bars (www.scotese.com/climate.htm) at the bottom represent time intervals during which global temperatures were as much as 10 °C warmer than today based on the climate reconstruction of Scotese; filled diamonds represent our estimates of carbonate fossil growth temperatures. Error bars represent ±1 s.e. on the Δ_{47} temperatures of well-preserved samples.

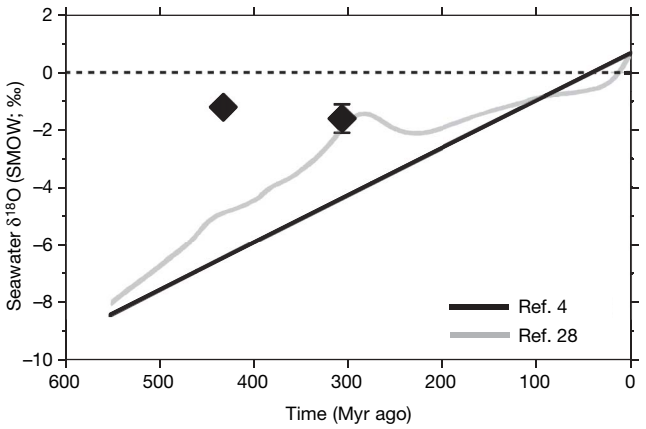


Figure 3 | Estimates of the oxygen isotopic composition of Phanerozoic sea water. The black[4] and grey[28] curves represent previous model estimates; the dashed line represents modern seawater $\delta^{18}O$ (SMOW); filled diamonds represent new estimates of Pennsylvanian and Silurian seawater $\delta^{18}O$ calculated from carbonate $\delta^{18}O$ and Δ_{47} temperatures. Error bars represent ±1 s.e. on the seawater $\delta^{18}O$ of well-preserved samples.

Our re-interpretation of the $\delta^{18}O$ values of Silurian and Pennsylvanian carbonate fossils also may apply to other parts of the Palaeozoic. However, there remain several marked discrepancies between climate reconstructions using the GEOCARB model versus those implied by the Scotese geological record and the Veizer *et al.* oxygen isotope record (which generally agree with each other, at least in timing of climate variations), and it is difficult to imagine that all time periods will be resolved in the same way as those examined in this study. For example, many carbonate fossils from the Cambrian and early Ordovician are so negative in $\delta^{18}O_{PDB}$ (in the range $-8‰$ to $-10‰$) that they cannot plausibly represent precipitation from an ocean with seawater $\delta^{18}O_{SMOW} \approx 0‰$, because in that case they would imply growth temperatures (54–67 °C) far in excess of the maximum temperature at which shallow-marine organisms can survive (37 °C)[26]. Application of carbonate clumped isotope thermometry to these extreme samples could reveal whether their low $\delta^{18}O$ values reflect consistently high levels of post-depositional alteration or low $\delta^{18}O$ values of sea water.

METHODS SUMMARY

CO_2 was extracted from all samples by phosphoric acid digestion using the laboratory methods described in ref. 5. Product CO_2 was analysed using a Finnigan MAT 253 gas source mass spectrometer configured to collect masses 44–49, inclusive, and standardized by comparison with CO_2 gases of known isotopic composition that had been heated for two hours at 1,000 °C to achieve a stochastic isotopic distribution[27] (see Methods). Several heated gas standards, spanning a range of bulk stable isotope compositions, were analysed to minimize the potential errors associated with mass spectrometric nonlinearities, which are observable when the compositions of samples and standards differ by more than 20–30‰ in any given isotope ratio[27] (see Methods). Masses 48 and 49 were monitored to assure adequate sample purification. Each measurement consisted of 6–9 acquisitions, with typical standard deviations (acquisition-to-acquisition) of 0.02‰ to 0.05‰ in Δ_{47} (see Table 1 footnote). Values of $\delta^{18}O$ and $\delta^{13}C$ were acquired as part of each analysis. Measured values of Δ_{47} were used to estimate carbonate growth temperature (T, in kelvin) using the relationship[5]:

$$\Delta_{47} = 0.0592(10^6\, T^{-2}) - 0.02.$$

Analyses of modern molluscs and brachiopods establish that this relationship holds for these forms of biogenic carbonate (see Methods). Paired temperature and carbonate $\delta^{18}O$ data were used to calculate the $\delta^{18}O$ value of formation and/or diagenetic waters using previously published calibrations of the temperature dependence of carbonate-water fractionations (see Methods).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 April; accepted 3 July 2007.

1. Berner, R. A. GEOCARBII: A revised model of atmospheric CO_2 over Phanerozoic time. *Am. J. Sci.* **294,** 56–91 (1994).
2. Berner, R. A. & Kothavala, Z. GEOCARBIII: A revised model of atmospheric CO_2 over Phanerozoic time. *Am. J. Sci.* **301,** 182–204 (2001).
3. François, L. M. & Walker, J. C. G. Modelling the Phanerozoic carbon cycle and climate: Constraints from the $^{87}Sr/^{86}Sr$ isotopic signature of seawater. *Am. J. Sci.* **292,** 81–135 (1992).
4. Veizer, J., Godderis, Y. & François, L. M. Evidence for decoupling of atmospheric CO_2 and global climate during the Phanerozoic eon. *Nature* **408,** 698–701 (2000).
5. Ghosh, P. *et al.* ^{13}C–^{18}O bonds in carbonate minerals: A new kind of paleothermometer. *Geochim. Cosmochim. Acta* **70,** 1439–1456 (2006).
6. Royer, D. L., Berner, R. A. & Park, J. Climate sensitivity constrained by CO_2 concentrations over the past 420 million years. *Nature* **446,** 530–532 (2007).
7. Ruddiman, W. F. *Earth's Climate: Past and Future* (Freeman, New York, 2001).
8. Montañez, I. P. *et al.* CO_2-forced climate and vegetation instability during Late Paleozoic deglaciation. *Science* **315,** 87–91 (2007).
9. Mora, C. I., Driese, S. G. & Colarusso, L. A. Middle to late Paleozoic atmospheric CO_2 levels from soil carbonate and organic matter. *Science* **271,** 1105–1107 (1996).
10. Yapp, C. J. & Poths, H. Ancient atmospheric CO_2 pressures inferred from natural goethites. *Nature* **355,** 342–344 (1992).
11. Crowley, T. J. & North, G. R. *Paleoclimatology* (Oxford Univ. Press, Oxford, 1991).
12. Caputo, M. V. & Crowell, J. C. Migration of glacial centers across Gondwana during Paleozoic Era. *Geol. Soc. Am. Bull.* **96,** 1020–1036 (1985).
13. Frakes, L. A. & Francis, J. E. A guide to Phanerozoic cold polar climates from high-latitude ice-rafting in the Cretaceous. *Nature* **333,** 547–549 (1988).
14. Boucot, A. J., Xu, C. & Scotese, C. R. Phanerozoic climate zones and paleogeography with a consideration of atmospheric CO_2 levels. *Paleont J.* **38,** 115–122 (2004).
15. Veizer, J. *et al.* $^{87}Sr/^{86}Sr$, $\delta^{13}C$ and $\delta^{18}O$ evolution of Phanerozoic seawater. *Chem. Geol.* **161,** 59–88 (1999).
16. Land, L. S. Comment on ''Oxygen and carbon isotopic composition of Ordovician brachiopods: Implications for coeval seawater'' by H. Qing and J. Veizer. *Geochim. Cosmochim. Acta* **59,** 2843–2844 (1995).
17. Azmy, K., Veizer, J., Bassett, M. G. & Copper, P. Oxygen and carbon isotopic composition of Silurian brachiopods: Implications for coeval seawater and glaciations. *Geol. Soc. Am. Bull.* **110,** 1499–1512 (1998).
18. Azmy, K., Veizer, J., Jin, J., Copper, P. & Brand, U. Paleobathymetry of a Silurian shelf based on brachiopod assemblages: An oxygen isotope test. *Can. J. Earth Sci.* **43,** 281–293 (2006).
19. Squires, R. L. *Burial Environment, Diagenesis, Mineralogy and Mg & Sr Contents of Skeletal Carbonates in the Buckhorn Asphalt of Middle Pennsylvanian Age, Arbuckle Mountains, Oklahoma.* PhD thesis, California Inst. Technol. (1973).
20. Brand, U. The oxygen and carbon isotopic composition of Carboniferous fossil components: Sea-water effects. *Sedimentology* **29,** 139–147 (1982).
21. Brand, U. Aragonite-calcite transformation based on Pennsylvanian molluscs. *Geol. Soc. Am. Bull.* **101,** 377–390 (1989).
22. Wenzel, B., Lécuyer, C. & Joachimski, M. M. Comparing oxygen isotope records of Silurian calcite and phosphate—$\delta^{18}O$ compositions of brachiopods and conodonts. *Geochim. Cosmochim. Acta* **64,** 1859–1872 (2000).
23. Gregory, R. T. & Taylor, H. P. An oxygen isotope profile in a section of Cretaceous oceanic crust, Samail ophiolite, Oman – Evidence for $\delta^{18}O$ buffering of the oceans by deep (less than 5 km) seawater-hydrothermal circulation at mid-ocean ridges. *J. Geophys. Res.* **86,** 2737–2755 (1981).
24. Muehlenbachs, K. The oxygen isotopic composition of the oceans, sediments and the seafloor. *Chem. Geol.* **145,** 263–273 (1998).
25. Kasting, J. F. *et al.* Paleoclimates, ocean depth, and the oxygen isotopic composition of seawater. *Earth Planet. Sci. Lett.* **252,** 82–93 (2006).
26. Brock, T. D. Life at high temperatures. *Science* **230,** 132–138 (1985).
27. Eiler, J. M. & Schauble, E. $^{18}O^{13}C^{16}O$ in Earth's atmosphere. *Geochim. Cosmochim. Acta* **68,** 4767–4777 (2004).
28. Wallmann, K. Impact of atmospheric CO_2 and galactic cosmic radiation on Phanerozoic climate change and the marine $\delta^{18}O$ record. *Geochem. Geophys. Geosys.* 5, Q06004 10.1029/2003GC000683 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank H. Affek, W. Guo and P. Ghosh for laboratory advice, and A. Wanamaker for assistance with samples.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.E.C. (rcame@gps.caltech.edu).

METHODS

Clumped isotope geochemistry. The carbonate 'clumped isotope' palaeothermometer involves the temperature-dependent 'clumping' of ^{13}C and ^{18}O (that is, the formation of bonds between these two rare isotopes) within the carbonate mineral lattice. The abundance of ^{13}C–^{18}O bonds in carbonate minerals depends on a thermodynamically controlled stable isotopic exchange equilibrium among various carbonate isotopologues, for example,:

$$Ca^{13}C^{16}O_3 + Ca^{12}C^{18}O^{16}O_2 = Ca^{13}C^{18}O^{16}O_2 + Ca^{12}C^{16}O_3 \quad (1)$$

Importantly, the temperature-dependent equilibrium constant for this reaction can be calculated on the basis of the isotopic composition of carbonate alone and does not require knowledge of the isotopic composition of the water in which the carbonate formed[5]. For this reason, carbonate clumped isotope thermometry can be applied rigorously to times and settings where the oxygen isotope composition of water is unknown.

The equilibrium constant for reaction (1) is reconstructed by isotopic analysis of CO_2 produced by reaction of carbonate with anhydrous phosphoric acid. These analyses involve simultaneous collection of ion beams corresponding to masses 44, 45, 46 (as for conventional measurements of the $\delta^{13}C$ and $\delta^{18}O$ values of carbonates) and 47. The mass-47 ion beam includes contributions from three isotopologues: $^{18}O^{13}C^{16}O$, $^{17}O^{12}C^{18}O$ and $^{17}O^{13}C^{17}O$. This population is overwhelmingly dominated by $^{18}O^{13}C^{16}O$, and so largely reflects the abundance of ^{13}C–^{18}O bonds in reactant carbonate. We define R^{47} as the ratio of the mass-47 isotopologues of CO_2 to the light isotopologue of CO_2 ($^{16}O^{12}C^{16}O$)[27]:

$$R^{47} = (\text{mass } 47)/(\text{mass } 44) \quad (2)$$

We report variations in R^{47} by comparison with the 'stochastic distribution'; that is, the relative abundance of isotopologues expected for a random distribution of all rare isotopes among all possible isotopologues. For a given set of $\delta^{18}O$, $\delta^{17}O$ and $\delta^{13}C$ values, R^{47} for the stochastic distribution is defined by[27]:

$$R^{47}{}_{\text{stochastic}} = (2[18][16][13] + [17]^2[13] + 2[18][17][12])/([16]^2[12]) \quad (3)$$

where [12] and [13] are the concentrations of ^{12}C and ^{13}C within the pool of all carbon atoms in the analysed CO_2, and [16], [17] and [18] are the concentrations of ^{16}O, ^{17}O and ^{18}O within the pool of all oxygen atoms in the analysed CO_2.

Finally, we use Δ_{47} to report how measured values of R^{47} differ from the stochastic distribution. Δ_{47} is defined as the difference in ‰ between the measured R^{47} value of the sample and the R^{47} value expected for that sample if its stable carbon and oxygen isotopes were randomly distributed among all isotopologues[27,29]:

$$\Delta_{47} = (R^{47}{}_{\text{measured}}/R^{47}{}_{\text{stochastic}} - 1) \times 1{,}000 \quad (4)$$

Laboratory methods. CO_2 was extracted from all samples using the laboratory methods described in ref. 5, which are an extension of well-established methods of phosphoric-acid digestion[30,31]. Product CO_2 was analysed using a Finnigan MAT 253 gas source mass spectrometer configured to collect masses 44–49 and standardized by comparison with CO_2 gases of known isotopic composition that had been heated for two hours at 1,000 °C to achieve a stochastic isotopic distribution[27]. Several heated gas standards, spanning a range of bulk stable isotope compositions, were analysed to minimize the potential errors associated with mass spectrometric nonlinearities, which are observable when the compositions of samples and standards differ by more than 20–30‰ in any given isotope ratio[27]. Masses 48 and 49 were monitored to assure adequate sample purification. Each measurement consisted of 6–9 acquisitions, with typical standard deviations (acquisition-to-acquisition) of 0.02‰ to 0.05‰ in Δ_{47}. As part of the analyses, $\delta^{18}O$ and $\delta^{13}C$ were simultaneously acquired. The Δ_{47} values were converted to carbonate growth temperature using the relationship[5]:

$$\Delta_{47} = 0.0592(10^6 T^{-2}) - 0.02 \quad (5)$$

Paired temperature and carbonate $\delta^{18}O$ data were used to calculate the $\delta^{18}O$ value of water from which the carbonates grew using the equation[32]:

$$10^3 \ln\alpha_{\text{calcite-water}} = 18.03(10^3 T^{-1}) - 32.42 + 0.25 \quad (6)$$

for calcitic brachiopods, and[33]:

$$10^3 \ln\alpha_{\text{aragonite-water}} = 18.45(10^3 T^{-1}) - 32.54 \quad (7)$$

for aragonitic molluscs, where α is the fractionation factor, T is temperature (in K) and $\delta^{18}O$ values are versus SMOW.

Appropriateness of the inorganic Δ_{47}-temperature calibration for brachiopods and molluscs. Ghosh *et al.*[5] determined the relationship between Δ_{47} and temperature (equation (5)) by analysing the CO_2 extracted from synthetic calcites grown in the laboratory at known, controlled temperatures. In addition, they analysed natural surface-dwelling corals (*Porites*) and deep-sea corals (*Desmophyllum dianthus*), which grew at known, approximately constant temperatures[5]. Their results indicate that the vital effects that influence the $\delta^{18}O$ and $\delta^{13}C$ of surface-dwelling and deep-sea corals do not influence the Δ_{47} values of the CO_2 extracted from that carbonate[5]. This result is consistent with models for vital effects[34], which describe the $\delta^{18}O$ and $\delta^{13}C$ offsets as reservoir effects, rather than kinetic fractionations. Recent calibration work[35] reveals that fish otolith carbonate, which also suffers from vital effects in $\delta^{18}O$ and $\delta^{13}C$, does not exhibit a significant offset from the Ghosh *et al.*[5] calibration of the relationship between growth temperature and the Δ_{47} of CO_2.

As part of the current study, we analysed naturally occurring brachiopods and molluscs (the two phyla from which we obtained Palaeozoic temperatures based on carbonate clumped isotope thermometry) that grew at known temperatures in the modern ocean (see Supplementary Table 1 and Supplementary Fig. 1). Our modern calibration materials agree very well (mean deviation of ±0.009‰ in Δ_{47} of CO_2) with the Ghosh *et al.*[5] temperature relationship for synthetic calcites.

Previous work on brachiopods and molluscs has shown that these organisms generally precipitate carbonate shells in isotopic equilibrium with the waters in which they form[36,37], without any apparent vital effects. Given this previous evidence for equilibrium carbonate growth in molluscs and brachiopods, and our new calibration results (see Supplementary Information), we suggest that vital effects do not influence the temperature estimates obtained for fossil molluscs and brachiopods based on carbonate clumped isotope thermometry.

29. Wang, Z., Schauble, E. A. & Eiler, J. M. Equilibrium thermodynamics of multiply substituted isotopologues of molecular gases. *Geochim. Cosmochim. Acta* **68**, 4779–4797 (2004).
30. McCrea, J. M. On the isotopic chemistry of carbonates and a paleotemperature scale. *J. Chem. Phys.* **18**, 849–857 (1950).
31. Swart, P. K., Burns, S. J. & Leder, J. J. Fractionation of the stable isotopes of oxygen and carbon in carbon dioxide during the reaction of calcite with phosphoric acid as a function of temperature and technique. *Chem. Geol.* **86**, 89–96 (1991).
32. Kim, S.-T. & O'Neil, J. R. Equilibrium and nonequilibrium oxygen isotope effects in synthetic carbonates. *Geochim. Cosmochim. Acta* **61**, 3461–3475 (1997).
33. Böhm, F. E. *et al.* Oxygen isotope fractionation in marine aragonite of coralline sponges. *Geochim. Cosmochim. Acta* **64**, 1695–1703 (2000).
34. Adkins, J. F., Boyle, E. A., Curry, W. B. & Lutringer, A. Stable isotopes in deep sea corals and a new mechanism for ''vital effect''. *Geochim. Cosmochim. Acta* **67**, 1129–1143 (2003).
35. Ghosh, P., Eiler, J. M., Campana, S. E. & Feeney, R. F. Calibration of the carbonate 'clumped isotope' paleothermometer for otoliths. *Geochim. Cosmochim. Acta* **71**, 2736–2744 (2007).
36. Brand, U., Logan, A., Hiller, N. & Richardson, J. Geochemistry of modern brachiopods: Applications and implications for oceanography and paleoceanography. *Chem. Geol.* **198**, 305–334 (2003).
37. Wanamaker, A. D. *et al.* An aquaculture-based method for calibrated bivalve isotope paleothermometry. *Geochem. Geophys. Geosyst.* **7**, Q09011, doi:10.1029/2005GC001189 (2006).

LETTERS

A link between large mantle melting events and continent growth seen in osmium isotopes

D. G. Pearson[1], S. W. Parman[1] & G. M. Nowell[1]

Although Earth's continental crust is thought to have been derived from the mantle, the timing and mode of crust formation have proven to be elusive issues. The area of preserved crust diminishes markedly with age[1,2], and this can be interpreted as being the result of either the progressive accumulation of new crust[3] or the tectonic recycling of old crust[4]. However, there is a disproportionate amount of crust of certain ages[1,2], with the main peaks being 1.2, 1.9, 2.7 and 3.3 billion years old; this has led to a third model in which the crust has grown through time in pulses[1,2,5–7], although peaks in continental crust ages could also record preferential preservation. The ^{187}Re–^{187}Os decay system is unique in its ability to track melt depletion events within the mantle and could therefore potentially link the crust and mantle differentiation records. Here we employ a laser ablation technique to analyse large numbers of osmium alloy grains to quantify the distribution of depletion ages in the Earth's upper mantle. Statistical analysis of these data, combined with other samples of the upper mantle, show that depletion ages are not evenly distributed but cluster in distinct periods, around 1.2, 1.9 and 2.7 billion years. These mantle depletion events coincide with peaks in the generation of continental crust and so provide evidence of coupled, global and pulsed mantle–crust differentiation, lending strong support to pulsed models of continental growth by means of large-scale mantle melting events[6].

The detailed timing of continental crust extraction should be recorded in the radiogenic isotope composition of the mantle. A systematically declining distribution of mantle depletion ages would support the steady accumulation model. Peaks in the depletion ages that corresponded to the zircon crustal age peaks would support the pulsed growth model.

The main impediment to seeing the record of crust extraction in the mantle is the disturbance of most radiogenic isotope systems by crustal recycling, metasomatic activity and dilution of the signal by convection. These processes re-enrich parts of the mantle and destroy the signature of melt depletion. Even in continental lithospheric mantle, isolated from convection for billions of years, Sr, Nd and Pb isotope signatures are dominated by enrichment processes[8]. We have therefore been unable to observe clearly the history of crust formation from the mantle.

The Re–Os isotope system has been singularly successful in tracing the depletion history of the subcontinental lithospheric mantle over 3-Gy periods despite equilibration temperatures of more than 1,000 °C (see, for example, refs 8, 9). This is due to its greater robustness to re-enrichment by melts and because osmium resides in dispersed trace phases that are less likely to re-equilibrate by diffusion. These properties also make the system useful for tracing the melt depletion history of the convecting mantle. Two types of sample useful for tracing depletion in the convecting mantle are abyssal peridotites, which increasingly seem to document melting events older than the age of the ridge they are dredged from[10,11], and osmium-rich platinum-group alloy grains (PGAs) derived from ophiolites, which also seem to record melting events older than the oceanic lithosphere in which they are found[12,13]. Here we use published abyssal and cratonic peridotite data along with published and new analyses of PGAs to investigate further the link between crust and mantle differentiation events.

We have made new osmium-isotope analyses of PGAs from the Urals, Tasmania and Tibet. We used a rapid but precise laser ablation ICPMS analytical technique (Supplementary Methods). All PGA grains studied were osmium-rich Os–Ir–Ru alloys[14].

The PGAs studied here form during chromite mineralization events in oceanic lithosphere that subsequently become ophiolites. Silicate-melt, chromite and sulphide inclusions within PGAs clearly indicate their magmatic formation[15]. The parental melts scavenge osmium from the mantle source and hence their osmium isotope compositions reflect source heterogeneities generated from previous melting events. Once formed, the low-Re/Os, osmium-rich PGAs are robust recorders of the source osmium isotope composition.

The four PGA locations studied range in formation age from 95 to 510 Myr. ^{187}Os/^{188}Os values vary widely, from 0.109 to 0.16 (Supplementary Information). The data from each location have probability density distributions indicative of multiple populations, with a main peak and one or more subsidiary peaks. The association of the PGAs with ophiolitic chromites ties the melting regime that transfers their osmium isotope signature from mantle to crust to a back-arc setting. The boninitic character of their trapped melts[15] confirm a previously depleted source and indicate PGAs forming from numerous melt infiltration events into oceanic lithosphere. Depleted ^{187}Os/^{186}Os values have not been not manifested in present-day mid-ocean-ridge and oceanic-island basalts analysed so far because their melt budget is strongly influenced by low-melting-point components such as pyroxenites[16] and high-Re/Os metasomatic sulphides[17]. Only in melting environments where these components have been largely melted out, and where melting of refractory, previously depleted mantle is assisted by the addition of water, can the depleted component dominate the osmium isotopic composition of some melts. Such water-rich melts have been shown to crystallize podiform chromites and hence to provide an environment for PGA formation[18] and the transfer of depleted osmium isotope signatures from the convecting mantle into the lithosphere.

To allow comparison with other records of mantle depletion, the PGA osmium isotope compositions can be recast as model ages, assuming the complete removal of rhenium from the source during melting. The resulting ages are direct functions of the model reservoir selected to simulate upper-mantle osmium isotope evolution. Extreme estimates for upper-mantle osmium isotope evolution models yield an uncertainty in the accuracy of any age of 0.4 Gyr but do not change the relative positions of age peaks (Supplementary Information). We use

[1]Northern Centre for Isotopic and Elemental Tracing, Department of Earth Sciences, Durham University, South Road, Durham DH1 4QE, UK.

ordinary (O) chondrites[19] as our reference mantle reservoir because this provides age estimates that are intermediate between the extremes produced by carbonaceous chondrites and primitive upper mantle.

Probability density plots of model ages (Fig. 1) show multiple peaks for all suites. The primary age peak for the PGAs and abyssal peridotites, samples of recent convecting upper mantle, is between 200 and 500 Myr, indicating evolution with an average Re/Os ratio slightly less than average O-chondrites. In the PGA suites the most prominent secondary model age peak is at 1.2 Gyr. This peak is also evident in the abyssal peridotite data and is prominent in the age distribution of cratonic peridotites from southern Africa and their sulphides. Given the differences in the physical nature of the PGA and peridotite/sulphide sample sets together with their global distribution, the degree of correspondence between the model age peaks recorded is remarkable. Another, smaller age peak is present in the Urals PGA data, abyssal peridotites and cratonic xenoliths at about 1.9 Gyr, and the Californian PGAs have two grains with ages of 2.7 Gyr. These ancient ages are clear evidence of the ability of the Re–Os isotope system to retain the imprint of ancient melting events, but our currently limited understanding of osmium diffusion in the convecting mantle does not yet permit a complete understanding of how the record is preserved.

To provide an independent estimate of the number and approximate ages of components present in our PGA data sets, we adopted a maximum-likelihood mixture modelling approach[20]. In all two component mixture models (Supplementary Methods), a 1.0–1.2-Gyr component is produced. When additional age components are introduced, the roughly 1.2-Gyr population persists and forms between 10% (California and Tibet) and 43% (Tasmania) of the population. Mixture modelling independently identifies components at about 1.6–1.9 Gyr in the abyssal, Urals and California data and also picks out a population at 2.6 Gyr within the California data in all models with more than two components. The use of a bayesian mixture modelling approach and Markov-chain Monte Carlo methods[21] produces very similar results.

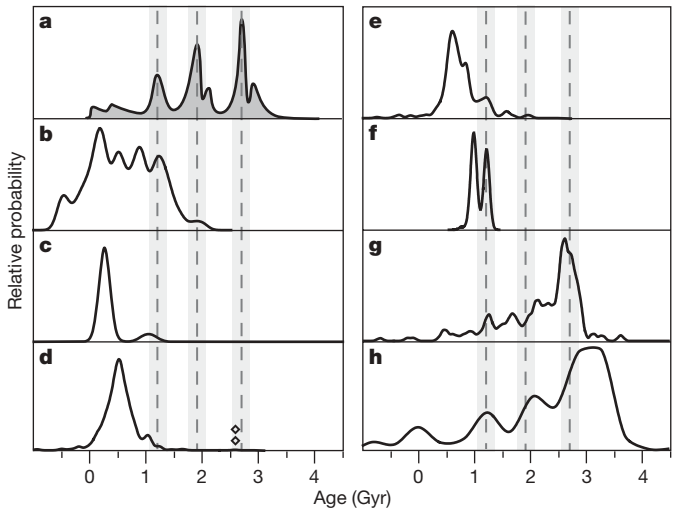


Figure 1 | Continental crust zircon ages compared with ages recorded in mantle samples. Probability density graphs of crustal zircon U-Pb ages[6] (**a**) plus osmium model ages (T_{RD}, Gyr) for mantle samples (**b**–**h**). Mantle samples comprise osmium-rich platinum-group metal alloys[12,13,24–27] (see Supplementary Information) (**c**–**f**), abyssal peridotites[10,11] (**b**), southern African cratonic peridotites[8,9] (**g**) and sulphides from cratonic peridotites[28,29] (**h**). Two stacked diamonds in **d** denote two data points at 2.6–2.7 Gyr. Ages are rhenium depletion ages (T_{RD}) for PGAs and peridotites but are calculated from measured Re/Os ratios (T_{MA}) for sulphides (see ref. 8 for details), to account for their significant Re/Os ratios. Ophiolite emplacement ages for PGAs together with numbers of samples for PGAs, southern African peridotites and their sulfides are as follows: **b**, abyssal, 0–50 Myr, $n = 80$; **c**, Tibet, 95 Myr, $n = 274$; **d**, California, 165 Myr, $n = 721$; **e**, Urals, 400 Myr, $n = 339$; **f**, Tasmania, 518 Myr, $n = 80$; **g**, $n = 228$; **h**, $n = 262$. Data for the Urals, Tasmania and Tibet are from this study and refs 24, 27 (see Supplementary Information). All data and probability density plots of osmium isotope ratios are given in Supplementary Information and were constructed with the program described in ref. 30. Graphs were plotted with uniform internal errors on individual model ages of 0.1 Gyr, to avoid overemphasis on single data points determined to high internal precision. The accuracy of the model ages is illustrated by the shaded zones surrounding the reference lines at 1.2, 1.9 and 2.7 Gyr and is taken from accuracy estimates using the different mantle reference reservoirs (see Supplementary Methods). Parameters for the calculation of model ages relative to the O-chondrite reservoir[19] are $^{187}Os/^{188}Os = 0.1283$ and $^{187}Re/^{188}Os = 0.422$.

It is important to evaluate the statistical significance of the age coincidences and to what extent the age distributions could be explained by heterogeneities in the mantle that have evolved from variable Re/Os and/or $^{187}Os/^{188}Os$ ratios in the original material that accreted to form the Earth. Of the chondritic meteorites analysed for osmium isotopes[19], one has an anomalously low $^{187}Os/^{188}Os$ value, yielding a model age equivalent to 1.2 Gyr. Could the model ages of mantle samples simply reflect original mantle osmium isotope heterogeneity? If so, a convecting mantle reservoir with a relatively low Re/Os ratio should still evolve isotopically through time (Fig. 2) because the anomalous chondrite sample has a Re/Os ratio only 27% lower than average. Any peaks in the isotopic composition of mantle rocks sampling this reservoir should also change position through time to reflect this evolution. This is clearly not true for the PGA and mantle data sets. Although the primary mode of each sample varies back through time, the secondary peak is stationary, at $^{187}Os/^{188}Os \approx 0.120$, equivalent to an age of 1.2 Gyr (Fig. 2). This indicates that the process forming the PGAs samples mantle regions containing additional, non-chondritic components of consistent isotopic composition that must have evolved with a Re/Os ratio close to zero. This almost invariant isotopic evolution will take place only if the source of this signature represents a residue from large degrees of melting, probably in excess of 35%, reducing the Re/Os ratio to very low values. These low-Re/Os sources must have remained in the convecting upper mantle for Gyr timescales, possibly trapped in areas of unusual mantle flow such as subduction zone wedges[22].

Monte Carlo simulations were performed to test the hypothesis that the data are single, normally distributed populations and that the secondary peaks are simply random fluctuations. No secondary 1.2-Gyr age-equivalent peaks were reproduced that were as large as those observed in the real data (Supplementary Information). Furthermore, no simulations produced any data with ages of 2.7 Gyr, indicating the very low probability that such ancient ages are random features. The likelihood of a specific peak's being repeated in different data sets was also tested by Monte Carlo simulation. Using even the widest likely normally distributed PGA data set (Fig. 3, thick lines) and assuming uniform probability, the chance of the California, Urals, Tasmania and Tibet data sets producing coincident age peaks at about 1.2 Gyr, in agreement with the zircon data, is 1 in 10^4. Using our preferred normal distribution, which matches the shapes of the major peaks in all the data sets, gives a probability of 1 in 10^{10} that the age correspondence in PGA data sets is fortuitous. Observing, by chance, only one PGA data set with a 2.7-Gyr age peak coincident with that of crustal zircons has a probability of less than 1 in 10^5. Therefore the coincidences between mantle Re–Os depletion ages and crustal zircon ages are highly significant.

The secondary peaks evident on a global scale seem to be recording large melting events superposed on a general trend of mantle depletion, indicating that an unusually large amount of depleted, low-Re/Os mantle formed at these times. These major melt depletion events, recorded in a diverse array of sample types over a wide geographic area, correspond very well to the crustal zircon age peaks (Fig. 1), so the zircon record must reflect crustal growth during major mantle melting events at these times. It is highly unlikely that such a coincidence could arise from preferential preservation of crust. The crust–mantle age correspondence provides very substantial support

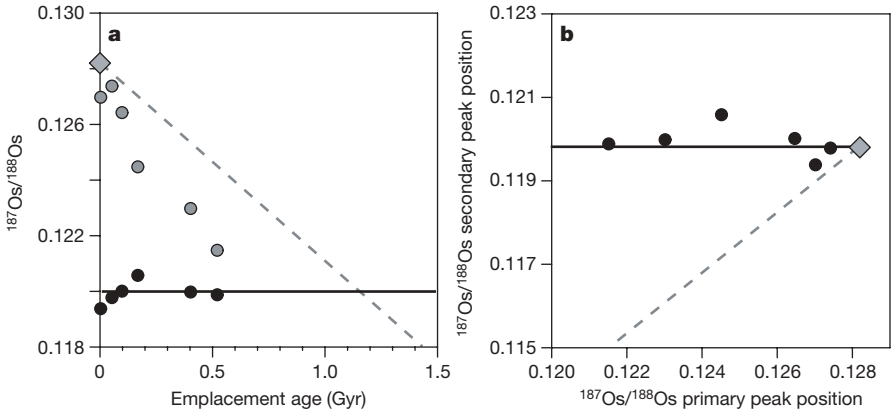


Figure 2 | Assessment of whether age peaks are due to mantle heterogeneity. a, $^{187}Os/^{188}Os$ ratio of the primary and secondary peaks for PGA suites, abyssal peridotites and peridotites from the Izu–Bonin–Mariana trench as a function of emplacement age. Peak positions were taken from probability density graphs and verified by mixture modelling (Supplementary Methods). Dashed line shows the evolution of average ordinary chondrite (parameters given in Fig. 1). Grey circles are $^{187}Os/^{188}Os$ values of the primary peak, which are always sub-chondritic; black circles give the position of the main secondary peaks. Emplacement ages are taken from Fig. 1. **b**, The relative positions of the primary and secondary age peaks. The position of the secondary peak in the PGA and abyssal peridotite data remains constant (horizontal line) as a result of a large-degree melt depletion event at 1.2 Gyr that decreases the Re/Os ratio to close to zero, whereas the position of the primary peak changes. Mantle heterogeneity of the type displayed by chondrites would result in data following the trajectory of the dashed grey line.

for crustal growth models that are punctuated by intense growth periods[5–7]. The periods of enhanced crust generation have been termed 'superevents' and can be can be recognized as pulses of formation of granite–greenstone terranes possibly linked to superplume events[6]. The pulsating nature of mantle depletion and crust formation revealed by osmium isotopes concurs with the picture of crust–mantle differentiation evident from convecting mantle helium-isotope systematics[23], in which peaks in helium isotopic compositions are interpreted to reflect the same major melt extraction events as those identified here.

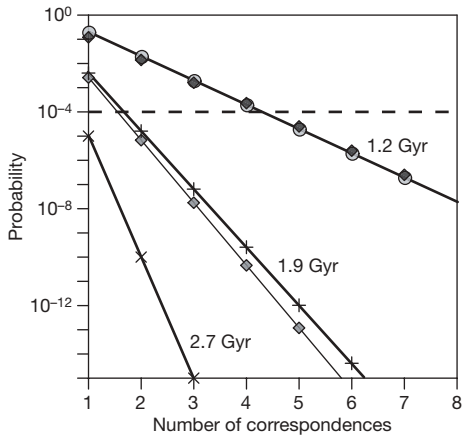


Figure 3 | Probability that the secondary age peaks (Fig. 1) match by random chance as the number of matching localities increases. Grey circles show probabilities calculated assuming the data are uniformly distributed, that is, a data point is equally likely to fall at any location within 99% of the range of the data. As all ages are equally likely, this curve applies to peaks of any age. All probabilities below 10^{-6} were extrapolated from the Monte Carlo modelling (see Supplementary Methods). When the abyssal peridotite data are included, the probability decreases from 1 in 10^4 (see the text) to 1 in 10^5. In fact, the data seem to be quasi-normally distributed (older ages less likely). Thus, the uniform distribution overestimates the probability of these older peaks and provides an upper limit to the probabilities. The use of a best-fit normal curve (primary position = 0.128, variance = 0.003; thick lines) yields very similar probabilities for the 1.2-Gyr peak (black diamonds), but much lower probabilities for the 1.9-Gyr (plus signs) and 2.7-Gyr (crosses) peaks, because data at older ages becomes increasingly unlikely with a normal distribution. Alternatively, the data can be viewed as containing two populations (a primary peak and a secondary peak). In this case, the primary peak represents the distribution of the data (primary position = 0.128, variance = 0.001). This is a much narrower distribution, and older ages are extremely unlikely. In this more realistic case, the chances of four data sets fortuitously giving an age peak correspondence at 1.2 Gyr (grey diamonds) is 1 in 10^{10}.

The zircon record in the continental crust retains abundant evidence of ancient events, whereas the number of ancient osmium model ages from the mantle declines sharply with increasing time (Fig. 1). The declining abundance of ancient depletion ages for PGA grains is a function of the mixing efficiency of mantle convection (Fig. 1). Although it is still possible to see evidence of the roughly 1.9-Gyr and 2.7-Gyr mantle depletion events, their clearer resolution and the detection of even older crust extraction events requires PGAs from older ophiolites than those sampled here.

The roughly 1.2-Gyr 'event' is the most prominent of the old mantle osmium model age peaks. However, this event is the least well defined in the zircon record in the continental crust and has been taken to indicate a less voluminous addition of juvenile material to the crust at 1.2 Gyr (refs 6, 7). However, the signature of major mantle melt extraction at this time is clear and widespread and could suggest unexposed additions to the lower crust that have not yet been documented, or the generation of dominantly zircon-poor basaltic crust at that time. Overall, the striking correlation between the timing of major crustal and mantle melting events is strongly supportive of the punctuated generation of juvenile continental material linked to major melting of the mantle.

Received 9 May; accepted 26 July 2007.

1. Gastil, G. The distribution of mineral dates in space and time. *Am. J. Sci.* **258**, 1–35 (1960).
2. Hurley, P. M. & Rand, J. R. Pre-drift continental nuclei. *Science* **164**, 1229–1242 (1969).
3. Allègre, C. J. & Rousseau, D. The growth of the continents through geological time studied by Nd isotope analysis of shales. *Earth Planet. Sci. Lett.* **67**, 19–34 (1984).
4. Armstrong, R. L. Radiogenic isotopes: the case for crustal recycling on a near-steady-state no-continental-growth Earth. *Phil. Trans. R. Soc. Lond. A* **301**, 443–472 (1981).
5. Goldstein, S. L., Arndt, N. T. & Stallard, R. F. The history of a continent from U–Pb ages of zircons from Orinoco River sand and Sm–Nd isotopes in Orinoco basin river sediments. *Chem. Geol.* **139**, 271–286 (1997).
6. Condie, K. C. Episodic continental growth and supercontinents. *Earth Planet. Sci. Lett.* **163**, 97–108 (1998).
7. Kemp, A. I. S., Hawkesworth, C. J., Paterson, B. A. & Kinny, P. D. Episodic growth of the Gondwana supercontinent from hafnium and oxygen isotopes in zircon. *Nature* **439**, 580–583 (2006).

8. Pearson, D. G. in *Mantle Petrology: Field Observations and High Pressure Experimentation* (eds Fei, Y., Bertka, C. M. & Mysen, B. O.) 57–78 (Spec. Pub. Geochem. Soc. no. 6, Houston, TX, 1999).
9. Carlson, R. W., Pearson, D. G. & James, D. E. Physical, chemical and chronological characteristics of continental mantle. *Rev. Geophys.* **43,** 1–24 (2005).
10. Brandon, A. D., Snow, J. E., Walker, R. J., Morgan, J. W. & Mock, T. D. ^{190}Pt–^{186}Os and ^{187}Re–^{187}Os systematics of abyssal peridotites. *Earth Planet. Sci. Lett.* **177,** 319–355 (2000).
11. Harvey, J. *et al.* Ancient melt extraction from the oceanic upper mantle revealed by Re–Os isotopes in abyssal peridotites from the Mid-Atlantic ridge. *Earth Planet. Sci. Lett.* **244,** 606–621 (2006).
12. Meibom, A. & Frei, R. Evidence for an ancient osmium isotopic reservoir in Earth. *Science* **296,** 516–518 (2002).
13. Meibom, A. *et al.* Re-Os isotopic evidence for long-lived heterogeneity and equilibration processes in the Earth's upper mantle. *Nature* **419,** 705–708 (2002).
14. Harris, D. C. & Cabri, L. J. Nomenclature of platinum-group-alloys: Review and revision. *Can. Mineral.* **29,** 231–237 (1991).
15. Brenker, F. E., Meibom, A. & Frei, R. On the formation of peridotite-derived Os-rich PGE alloys. *Am. Mineral.* **88,** 1731–1740 (2003).
16. Sobolev, A. V. *et al.* The amount of recycled crust in sources of mantle-derived melts. *Science* **316,** 412–417 (2007).
17. Alard, O. *et al. In situ* Os isotopes in abyssal peridotites bridge the isotopic gap between MORBs and their source mantle. *Nature* **436,** 1005–1008 (2005).
18. Mateev, S. & Balhaus, C. Role of water in the origin of podiform chromitite deposits. *Earth Planet. Sci. Lett.* **203,** 235–243 (2002).
19. Walker, R. J. *et al.* Comparative ^{187}Re–^{187}Os systematics of chondrites: implications regarding early solar system processes. *Geochim. Cosmochim. Acta* **66,** 4187–4201 (2002).
20. Sambridge, M. S. & Compston, W. Mixture modelling of multi-component data sets with application to ion-probe zircon ages. *Earth Planet. Sci. Lett.* **128,** 373–390 (1994).
21. Jasra, A., Stephens, D. A., Gallagher, K. & Holmes, C. C. Bayesian mixture modelling in geochronology via Markov chain Monte Carlo. *Math. Geol.* **38,** 269–300 (2006).
22. Parkinson, I. J., Hawkesworth, C. J. & Cohen, A. S. Ancient mantle in a modern arc: Osmium isotopes in Izu–Bonin–Mariana forearc peridotites. *Science* **281,** 2011–2013 (1998).
23. Parman, S. W. Helium isotopic evidence for episodic mantle melting and crustal growth. *Nature* **446,** 900–903 (2007).
24. Hattori, K. & Hart, S. R. Osmium-isotope ratios of platinum-group minerals associated with ultramafic intrusions; Os-isotopic evolution of the oceanic mantle. *Earth Planet. Sci. Lett.* **107,** 499–514 (1991).
25. Walker, R. J. *et al.* ^{187}Os–^{186}Os systematics of Os–Ir–Ru alloy grains from southwestern Oregon. *Earth Planet. Sci. Lett.* **230,** 211–226 (2005).
26. Brandon, A. D., Walker, R. J. & Puchtel, I. Platinum–osmium isotope evolution of the Earth's mantle: Constraints from chondrites and Os-rich alloys. *Geochim. Cosmochim. Acta* **70,** 2093–2103 (2006).
27. Shi, R. D. *et al.* Multiple events in the Neo-Tethyan oceanic upper mantle: evidence from Ru–Os–Ir alloys in the Luobusa and Dongqiao ophiolitic podiform chromitites, Tibet. *Earth Planet. Sci. Lett.*, doi:10.1016/j.epsl.2007.05.044 (2007).
28. Griffin, W. L., Graham, S., O'Reilly, S. Y. & Pearson, N. J. Lithosphere evolution beneath the Kaapvaal Craton: Re–Os systematics of sulfides in mantle-derived peridotites. *Chem. Geol.* **208,** 89–118 (2004).
29. Griffin, W. L., Spetsius, Z. V., Pearson, N. J. & O'Reilly, S. Y. In situ Re–Os analysis of sulfide inclusions in kimberlitic olivine: New constraints on depletion events in the Siberian lithosphere. *Geochem. Geophys. Geosyst.* **3,** 1069, doi:10.1029/2001GC000287 (2002).
30. Ludwig, K. R. Isoplot. Program and documentation, version 2.95. Revised edition of US Open-File report 91–445 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank P. Nixon, the British Museum of Natural History, C. Francis of the Harvard Museum and the Tasmanian Geological Survey for the supply of PGAs used in this study, A. Brandon for making the paper more robust, L. Jaques for advice on sourcing PGAs, and M. Goldstein and K. Gallagher for guidance on statistical approaches.

Author Contributions All authors contributed equally to this study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.G.P. (d.g.pearson@durham.ac.uk).

LETTERS

Placing late Neanderthals in a climatic context

P. C. Tzedakis[1,2], K. A. Hughen[3], I. Cacho[4] & K. Harvati[5]

Attempts to place Palaeolithic finds within a precise climatic framework are complicated by both uncertainty over the radiocarbon calibration beyond about 21,500 ^{14}C years BP[1] and the absence of a master calendar chronology for climate events from reference archives such as Greenland ice cores or speleothems[2]. Here we present an alternative approach, in which ^{14}C dates of interest are mapped directly onto the palaeoclimate record of the Cariaco Basin by means of its ^{14}C series[3], circumventing calendar age model and correlation uncertainties, and placing dated events in the millennial-scale climate context of the last glacial period. This is applied to different sets of dates from levels with Mousterian artefacts, presumably produced by late Neanderthals, from Gorham's Cave in Gibraltar: first, generally accepted estimates of about 32,000 ^{14}C years BP for the uppermost Mousterian levels[4,5]; second, a possible extended Middle Palaeolithic occupation until about 28,000 ^{14}C years BP[6]; and third, more contentious evidence for persistence until about 24,000 ^{14}C years BP[6]. This study shows that the three sets translate to different scenarios on the role of climate in Neanderthal extinction. The first two correspond to intervals of general climatic instability between stadials and interstadials that characterized most of the Middle Pleniglacial and are not coeval with Heinrich Events. In contrast, if accepted, the youngest date indicates that late Neanderthals may have persisted up to the onset of a major environmental shift, which included an expansion in global ice volume and an increased latitudinal temperature gradient. More generally, our radiocarbon climatostratigraphic approach can be applied to any 'snapshot' date from discontinuous records in a variety of deposits and can become a powerful tool in evaluating the climatic signature of critical intervals in Late Pleistocene human evolution.

The Neanderthal extinction, and the possible implication of climate therein, has long been the subject of intense debate[7]. Current evidence points to southern Iberia as one of the last Neanderthal strongholds in Europe, where several Middle Palaeolithic occupation sites suggest the regional survival of Neanderthals up to about 30,000–32,000 ^{14}C yr BP[8]. Of these, Gorham's Cave, Gibraltar, has yielded both Middle and Upper Palaeolithic lithic assemblages, with the youngest dates from Mousterian layers converging on about 32,000 ^{14}C yr BP[5,6]. Recently, Finlayson *et al.*[6] provided 22 additional dates on charcoal from a newly excavated level with Mousterian artefacts deep within Gorham's Cave. Of these, the three lowermost dates (28,170 ± 240, 29,210 ± 190 and 32,560 ± 390 ^{14}C yr BP; all errors are reported at one standard deviation) were in stratigraphic order, but several age reversals occurred further up. This was attributed to repeated use (for example, trampling and cleaning) of the same location within the cave as a hearth site over several thousand years[6]. Within the presumed core of this recurrently used combustion area, three dates in stratigraphic order (24,010 ± 160, 26,400 ± 220 and 30,560 ± 360 ^{14}C yr BP) were considered to come from *in situ* superimposed hearths[6]. Although the Finlayson *et al.* results have come under scrutiny[9,10], taken at face value they suggest that Neanderthals may have survived in the area until about 28,000 and possibly until about 24,000 ^{14}C yr BP[6], thus extending the range for their last regional appearance by up to 6,000 ^{14}C yr. More recently, the youngest of these dates has been used to link[11] the final disappearance of Neanderthals from southern Iberia to the climatic conditions associated with Heinrich Event 2 (H2). In that model, the extreme conditions of H2 are proposed to have acted as a 'territory cleanser', allowing subsequent colonization by anatomically modern humans[11]. Over the years a substantial body of evidence from marine and terrestrial reference archives[12–16] and climate models[17] has demonstrated the coupling of southern Iberian environments to North Atlantic variability on millennial and orbital timescales, with the coldest and driest conditions occurring during Heinrich Events, whereas intervening Dansgaard/Oeschger Stadials (or Greenland Stadials, following the terminology of ref. 18) were less extreme. However, the complex sedimentation patterns of cave deposits means that the linking of the archaeological and climate records is not straightforward. Although animal and plant remains from the excavated levels provide some indication of the palaeoenvironmental situation around Gorham's Cave[6], they are discrete samples, representing wide time windows and not forming part of a long, highly resolved and continuous time series that could be compared and correlated with other palaeoclimate records. Moreover, the uncertainty of ^{14}C calibration[1] and differences, sometimes on the order of hundreds of years, between published timescales of reference climate archives[2] complicates the placement of the putative late Neanderthal occupation within the context of millennial-scale variability of the last glacial period. Thus, even if radiocarbon dates could be converted perfectly to calendar years, it would still be unclear where precisely they fit into the succession of rapid climate events of the last glacial period, complicating a proper evaluation of the role of climate in Neanderthal extinction.

One way around this problem is based on the observation that a well-constrained stratigraphic relationship between ^{14}C dates and climate can be established in certain reference sequences that contain a North Atlantic signal of millennial-scale variability. Ocean Drilling Program sediment cores from the Cariaco Basin, Venezuela, provide such a sequence, combining high-resolution palaeoclimate records with ^{14}C dates from planktonic foraminiferans. Previously, ^{14}C and palaeoclimate records from Cariaco ODP Site 1002 were linked to the calendar chronology of the layer-counted GISP2 Greenland ice core[19] to provide calendar age estimates of the ^{14}C timescale back to 50,000 yr BP[3]. However, linking ^{14}C dates from other sites to a Greenland climatic context by using such a comparison involves considerable uncertainty from multiple sources, including the correlation procedure and calendar age chronology itself. An alternative approach is to map ^{14}C dates of interest directly onto the climate record by using the Cariaco ^{14}C series and palaeoclimate series in the depth domain

[1]Earth and Biosphere Institute, School of Geography, University of Leeds, Leeds LS2 9JT, UK. [2]Department of Environment, University of the Aegean, Mytilene 81100, Greece. [3]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. [4]GRC Marine Geosciences, University of Barcelona, E-08028 Barcelona, Spain. [5]Department of Human Evolution, Max-Planck Institute for Evolutionary Anthropology, D-04103 Leipzig, Germany.

(Fig. 1; see Supplementary Table 1 for details). A stratigraphic comparison within the same sediment archive eliminates uncertainties arising from the conversion to calendar ages, and places ^{14}C ages into the climatic framework of the last glacial period.

The different sets of dates from Gorham's Cave discussed above represent an excellent testing ground for the radiocarbon climatostratigraphic approach outlined above. In essence they can be regarded as providing three distinct case studies on the timing of the latest Middle Palaeolithic: a set of generally accepted but older age determinations[4,5], younger dates suggesting a possible extended Neanderthal survival[6], and finally more contentious dates for late persistence[6]. Here we use the youngest date considered to be stratigraphically reliable within each data set: first, 32,330 ± 390 ^{14}C yr BP on charcoal identified as pine-cone scale, which is also statistically identical with several other dates from late Middle Palaeolithic levels at Gorham's Cave[4,5,10]; second, 28,170 ± 240 ^{14}C yr BP on charcoal from the lowermost three dates in stratigraphic order of Finlayson *et al.*[6]; and third, 24,010 ± 160 ^{14}C yr BP on charcoal from the presumed *in situ* superimposed hearths[6].

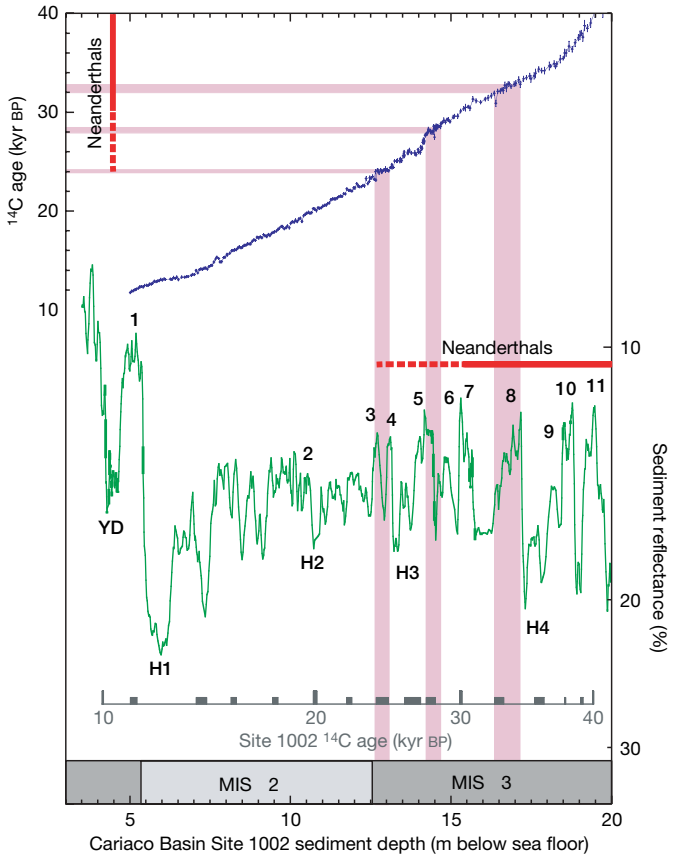


Figure 1 | Mapping Gorham's Cave ^{14}C dates onto the palaeoclimate record of the Cariaco Basin. Radiocarbon and palaeoclimate time series from Cariaco Basin ODP Site 1002 are plotted against depth. Cariaco marine ^{14}C ages (blue) were corrected for surface ocean reservoir age by subtracting 420 years before plotting. Potential changes in past reservoir age result in an additional ±100 years uncertainty[25], which is combined with reported analytical uncertainties according to established procedures[1,26] and plotted as 1σ errors. Sediment reflectance at 550 nm wavelength (green) is a proxy for organic carbon content and corresponds to upwelling, trade-wind intensity and nutrient runoff. Millennial-scale climate oscillations are numbered according to the corresponding Greenland Interstadials (GIS), and Heinrich Events (H) and the Younger Dryas (YD) are also indicated. Approximate ^{14}C ages for the Cariaco palaeoclimate record, taking into account changes in the ^{14}C chronology such as plateaux, are shown along the inset horizontal axis. The width of the tick marks reflects 1σ uncertainty; tick marks falling on a radiocarbon plateau therefore appear wider. Gorham's Cave ^{14}C dates are shown in pink and errors are reported as 1σ, following radiocarbon convention[27]. The dates are as follows: 32,330 ± 390 ^{14}C yr BP on charcoal from a pine-cone scale[5], considered stratigraphically reliable[10] and statistically identical with several other dates from late Middle Palaeolithic levels at Gorham's Cave[4,5]; 28,170 ± 240 ^{14}C yr BP on charcoal, the youngest of the lowermost three dates in stratigraphic order of Finlayson *et al.*[6]; and 24,010 ± 160 ^{14}C yr BP on charcoal, the youngest of the presumed *in situ* hearths[6]. Where the dates intersect the Cariaco ^{14}C chronology, depths can be translated into the sediment reflectance record for precise climatic context. Red bars indicate the late Neanderthal timeline and include the consensus interval for regional survival of Middle Palaeolithic Neanderthals[8] (solid) and their possible extended range[6] (dashed).

Figure 1 shows that the date of 32,330 ± 390 ^{14}C yr BP translates to an interval in the Cariaco record corresponding to Greenland Interstadial (GIS) 8 and the onset of the ensuing Greenland Stadial (GS) 8. Moreover, visual inspection reveals that the consensus age range of about 30,000–32,000 ^{14}C yr BP for the regional survival of late Neanderthals in southern Iberia[8] extends from the onset of GIS8 to the end of GIS7, an interval characterized by conditions typical of the climate instability of Marine Isotope Stage (MIS) 3 (that is, oscillations between Greenland Stadials and Interstadials). Similarly, the date of 28,170 ± 240 ^{14}C yr BP translates to an interval from the onset of GS6 to the middle of GIS5. None of the above dates falls within intervals corresponding to the more extreme events H3 and H4. In contrast, the younger date of 24,010 ± 160 ^{14}C yr BP extends the Neanderthal occupation to an interval immediately preceding the onset of a major shift in the dominant climate state, which marks the boundary between MIS3 and MIS2. The date falls on a radiocarbon plateau (a short interval of calendar age during which radiocarbon ages change slowly or not at all), which in the Cariaco record represents the interval corresponding to GIS4, GS4 and GIS3. Of particular interest is the end of GIS3, which marks the onset of a long period with reduced range of climate instability, representing the run to Last Glacial Maximum conditions with an accelerated decrease in sea level of 20–30 m (ref. 20) or more[21], and a corresponding expansion of land ice between GIS3 and H2 (Supplementary Fig. 1). In addition, the end of GIS3 marks a partial decoupling between high and low/middle latitudes: whereas Greenland ice cores and subpolar North Atlantic records show a return to cold conditions and increased ice rafting[22–24], the tropical and subtropical North Atlantic, including the Portuguese margin (Supplementary Fig. 1), show sea surface temperatures that continued to be relatively warm until H2. The persistence of interstadial conditions during this interval is also observed in the western Mediterranean Sea, where relatively high sea surface temperatures[12] are consistent with the entrance of warm water from the subtropical gyre, and also further east as suggested by the continued presence of temperate tree populations in Greece (Supplementary Fig. 1). The increased isolation of the Mediterranean Basin as a result of lower sea level may have amplified the warm conditions, resulting in the persistence of a relatively warm western Mediterranean deep water mass and increased precipitation and high lake levels in the eastern Mediterranean, with all lakes rising after 24,000 ^{14}C yr BP and reaching maximum levels between GIS3 and H2 (Supplementary Fig. 1).

Our analysis suggests that the consensus age range of about 30,000–32,000 ^{14}C yr BP for late Neanderthals in southern Iberia[4,5,8], as well as the less contentious age estimate of about 28,000 ^{14}C yr BP by Finlayson *et al.*[6], place their disappearance somewhere between GIS8 and GIS5. The character of this interval is not particularly distinct from the general climate instability of much of MIS3 and does not include Heinrich Events, thereby suggesting a limited role of climate in Neanderthal extinction. By comparison, taken at face value, the youngest and more contentious date of Finlayson *et al.*[6] suggests that Neanderthals at Gorham's Cave may have persisted up to the onset of a major environmental shift, which included an expansion in global ice volume and an increased latitudinal temperature gradient. This would imply a greater role of climate in Neanderthal extinction, not necessarily directly but perhaps in the form of climate-driven intensified competition as a result of

increased southward human migration from higher latitudes. This exercise shows that the three sets of dates translate to two very different scenarios on the role of climate in Neanderthal extinction; resolving the chronology of the site through further dating campaigns (as suggested in ref. 9) can therefore have important implications. Finally, with regard to the role of the climatic conditions associated with H2 in the Neanderthal demise[11], Fig. 1 and Supplementary Fig. 1 together clearly indicate that H2 postdates the youngest proposed Neanderthal date[6] by about 3,000 years. This provides an excellent illustration of how our radiocarbon climatostratigraphic approach can reduce the risks of conflating events that are chronologically distinct and therefore not causally related. What emerges is that despite current uncertainties over radiocarbon calibration and the absolute timing of widespread abrupt climate changes, this approach can be used to place critical events in the history of late archaic and modern humans into a precise palaeoclimatic context, leading to better-constrained scenarios on the underlying causes of the observed patterns.

Received 18 February; accepted 26 July 2007.

1. Reimer, P. J. *et al.* IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* **46,** 1029–1058 (2004).
2. Svensson, A. *et al.* The Greenland Ice Core Chronology 2005, 15–42 ka. Part 2: comparison to other records. *Quat. Sci. Rev.* **25,** 3258–3267 (2006).
3. Hughen, K. *et al.* C-14 activity and global carbon cycle changes over the past 50,000 years. *Science* **303,** 202–207 (2004).
4. Pettitt, P. B. & Bailey, R. M. in *Neanderthals on the Edge* (eds Stringer, C. B., Barton, R. N. E. & Finlayson, C.) 155–162 (Oxbow Books, Oxford, 2000).
5. Bronk Ramsey, C., Higham, T. F. G., Owen, D. C., Pike, A. W. G. & Hedges, R. E. M. Radiocarbon dates from the Oxford AMS System. Archaeometry Datelist 31. *Archaeometry* **44,** 1–149 (2002).
6. Finlayson, C. *et al.* Late survival of Neanderthals at the southernmost extreme of Europe. *Nature* **443,** 850–853 (2006).
7. Stringer, C. *et al.* in *Neanderthals and Modern Humans in the European Landscape during the Last Deglaciation* (eds van Andel, T. H. & Davies, W.) 233–240 (McDonald Institute Monographs, Cambridge, 2003).
8. Zilhão, J. Chronostratigraphy of the Middle-to-Upper Paleolithic transition in the Iberian peninsula. *Pyrenae* **37,** 7–84 (2006).
9. Delson, E. & Harvati, K. Return of the last Neanderthal. *Nature* **443,** 762–763 (2006).
10. Zilhão, J. & Pettitt, P. On the new dates for Gorham's Cave and the late survival of Iberian Neanderthals. *Before Farming* [online version] 2006/3 article 3 ⟨http://www.waspress.co.uk/journals/beforefarming/journal_20063/abstracts/index.php⟩ (2006).
11. Jiménez-Espejo, F. J. *et al.* Climate forcing and Neanderthal extinction in Southern Iberia: insights from a multiproxy marine record. *Quat. Sci. Rev.* **26,** 836–852 (2007).
12. Cacho, I. *et al.* Dansgaard–Oeschger and Heinrich event imprints in Alboran Sea paleotemperatures. *Paleoceanography* **14,** 698–705 (1999).
13. Moreno, A. *et al.* Links between marine and atmospheric processes oscillating at millennial time-scale. A multi-proxy study of the last 50,000 yr from the Alboran Sea (Western Mediterranean Sea). *Quat. Sci. Rev.* **24,** 1623–1636 (2005).
14. Pons, A. & Reille, M. The Holocene- and Upper Pleistocene pollen record from Padul (Granada, Spain), a new study. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **66,** 243–263 (1988).
15. Sánchez Goñi, M. F. *et al.* Synchroneity between marine and terrestrial responses to millennial scale climatic variability during the last glacial period in the Mediterranean region. *Clim. Dyn.* **19,** 95–105 (2002).
16. Combourieu-Nebout, N. *et al.* Enhanced aridity and atmospheric high-pressure stability over the western Mediterranean during the North Atlantic cold events of the past 50 k.y. *Geology* **30,** 863–866 (2002).
17. Sepulchre, P. *et al.* H4 abrupt event and late Neanderthal presence in Iberia. *Earth Planet. Sci. Lett.* **258,** 283–292 (2007).
18. Walker, M. J. C. *et al.* Isotopic 'events' in the GRIP ice core: a stratotype for the Late Pleistocene. *Quat. Sci. Rev.* **18,** 1143–1150 (1999).
19. Meese, D. A. *et al.* The Greenland Ice Sheet Project 2 depth-age scale: Methods and results. *J. Geophys. Res.* **102,** 26411–26423 (1997).
20. Siddall, M. *et al.* Sea-level fluctuations during the last glacial cycle. *Nature* **423,** 853–858 (2003).
21. Peltier, W. R. & Fairbanks, R. G. Global glacial ice volume and Last Glacial Maximum duration from and extended Barbados sea level record. *Quat. Sci. Rev.* **25,** 3322–3337 (2006).
22. Stuiver, M. & Grootes, P. M. GISP2 oxygen isotope ratios. *Quat. Res.* **53,** 277–283 (2000).
23. Weinelt, M. *et al.* Variability of North Atlantic heat transfer during MIS 2. *Paleoceanography* **18,** doi:10.1029/2002PA000772 (2003).
24. Hemming, S. R. Heinrich events: Massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42,** RG1005, doi:10.1029/2003RG000128 (2004).
25. Kromer, B. *et al.* Late Glacial ^{14}C ages from a floating 1270-ring pine chronology. *Radiocarbon* **46,** 1203–1210 (2004).
26. Hughen, K. A. *et al.* MARINE04 marine radiocarbon age calibration, 26–0 ka BP. *Radiocarbon* **46,** 1059–1086 (2004).
27. Stuiver, M. & Polach, H. A. Discussion: reporting of ^{14}C data. *Radiocarbon* **19,** 355–363 (1977).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank N. Galanidou and R. Preece for discussions, and E. Bard and E. Rohling for providing published data. We acknowledge a Fellowship from The Leverhulme Trust during 2006–2007 (P.C.T.) and support from the US NSF and S. M. Tudor (K.A.H.), the Comer Science and Education Foundation (USA) and the Ramón y Cajal programme of the Spanish MEC (I.C.), and the Max Planck Gesellschaft and the 'EVAN' Marie Curie Research Training Network (K.H.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.C.T. (p.c.tzedakis@leeds.ac.uk).

LETTERS

Positive feedbacks promote power-law clustering of Kalahari vegetation

Todd M. Scanlon[1], Kelly K. Caylor[2], Simon A. Levin[3] & Ignacio Rodriguez-Iturbe[4]

The concept of local-scale interactions driving large-scale pattern formation has been supported by numerical simulations, which have demonstrated that simple rules of interaction are capable of reproducing patterns observed in nature[1,2]. These models of self-organization suggest that characteristic patterns should exist across a broad range of environmental conditions provided that local interactions do indeed dominate the development of community structure. Readily available observations that could be used to support these theoretical expectations, however, have lacked sufficient spatial extent or the necessary diversity of environmental conditions to confirm the model predictions. We use high-resolution satellite imagery to document the prevalence of self-organized vegetation patterns across a regional rainfall gradient in southern Africa, where percent tree cover ranges from 65% to 4%. Through the application of a cellular automata model, we find that the observed power-law distributions of tree canopy cluster sizes can arise from the interacting effects of global-scale resource constraints (that is, water availability) and local-scale facilitation. Positive local feedbacks result in power-law distributions without entailing threshold behaviour commonly associated with criticality. Our observations provide a framework for integrating a diverse suite of previous studies that have addressed either mean wet season rainfall or landscape-scale soil moisture variability as controls on the structural dynamics of arid and semi-arid ecosystems.

Scale is an essential factor in linking pattern and process[3], and an adequate characterization of tree canopy distributions must span scales ranging from that of individual to that of the landscape. Large-scale plot studies such as those at Barro Colorado Island and a limited number of other locations have proven to be extremely valuable for defining vegetation characteristics such as canopy gap distribution[1,4] and species-specific clustering[5]. The massive amount of manual sampling required to compile these data sets, however, places a practical limitation on their widespread collection. High-resolution remote sensing is an alternative for detecting landscape-level vegetation pattern, and one that is particularly well suited for monitoring sparse vegetation in which individual tree canopies can be distinguished[6]. The ease by which these data can be acquired allows for a more geographically widespread detection of large-scale spatial patterns.

Inferring process from vegetation pattern has been a fundamental motivation of many landscape ecological studies, yet unambiguous determination of the factors that generate and maintain patterns is often obfuscated by the existence of multiple mechanisms that can give rise to commonly observed spatial arrangements. For example, random patterns can be indicative of the absence of spatial interactions, but random patterns are also known to emerge from strong competitive interactions[7]. Aggregated patterns, such as those routinely observed for woody tree species in natural communities, have been attributed to the disparate mechanisms of dispersal limitation[8] and habitat differentiation[9]. It has been suggested that the identification of dominant processes that lead to emergent vegetation pattern could benefit from a more thorough statistical measure of the vegetation spatial structure that implicitly considers the broad range of spatial scales over which aggregation occurs, rather than simply characterizes the average aggregation tendency of individuals[10]. We adopt such an approach in applying cluster size analysis to tree canopy distributions and evaluate the consistency of the patterns over a range of environmental conditions.

Remote sensing analysis focuses on the Kalahari Transect in southern Africa, one in a number of worldwide International Geosphere-Biosphere Programme transects[11] that spans a mean annual rainfall gradient from approximately 1,200 to 200 mm per year. IKONOS satellite images were acquired for six locations along the Kalahari Transect during the 2000 wet season to augment concurrent field surveys, which have produced detailed information about the savanna vegetation[12,13]. Common to each of these locations is the homogeneous sand formation that underlies them (Supplementary Fig. 1), a feature that provides constant background spectral reflectance for remote sensing and acts as a normalizing factor for comparing rainfall–vegetation relationships between sites. A strong correlation ($r^2 = 0.84$, $n = 10$) exists between mean wet season rainfall ($\bar{r}$) and fractional tree cover (f_t), both established from ground-based measurements[12]. Key descriptors of the individual sites are listed in Table 1 and spatial arrangements of tree canopies derived from the IKONOS images are shown in Fig. 1a.

We describe the distribution of cluster sizes within each landscape using the inverse cumulative distribution, which is the probability that a cluster area (A) is greater than or equal to a, $P(A \geq a)$. To

Table 1 | Kalahari site characteristics

Site	Lat./Lon.	$\bar{r}$ (mm)	f_t	$\bar{c}$ (m^2)	Dominant tree species
Mongu, Zambia	14.42° S, 23.52° E	879	0.65	14.4	*Brachystegia spiciformis*
Senanga, Zambia	15.86° S, 23.34° E	811	0.54	23.1	*Brachystegia spiciformis*
Pandamatanga, Botswana	18.66° S, 25.50° E	698	0.32	15.8	*Schinziophyton rautanenii*, *Baikiaea plurijuga*
Sandveld, Namibia	22.02° S, 19.17° E	409	0.19	3.3	*Terminalia sericea*
Tshane, Botswana	24.17° S, 21.89° E	365	0.14	10.3	*Acacia erioloba*
Vastrap, South Africa	27.75° S, 21.42° E	216	0.04	2.0	*Acacia erioloba*

Mean annual rainfall, $\bar{r}$, is extrapolated from nearby meteorological stations. Fractional tree cover, f_t, mean canopy area, $\bar{c}$, and the dominant species are derived from 0.25–1.0 ha stem map surveys of the sites[12,13]. Lat., latitude; Lon., longitude; S, south; E, east.

[1]Department of Environmental Sciences, University of Virginia Charlottesville, Virginia 22903, USA. [2]Department of Geography, Indiana University Bloomington, Indiana 47401, USA. [3]Department of Ecology & Evolutionary Biology, Princeton University Princeton, New Jersey 08544, USA. [4]Department of Civil and Environmental Engineering, Princeton University Princeton, New Jersey 08544, USA.

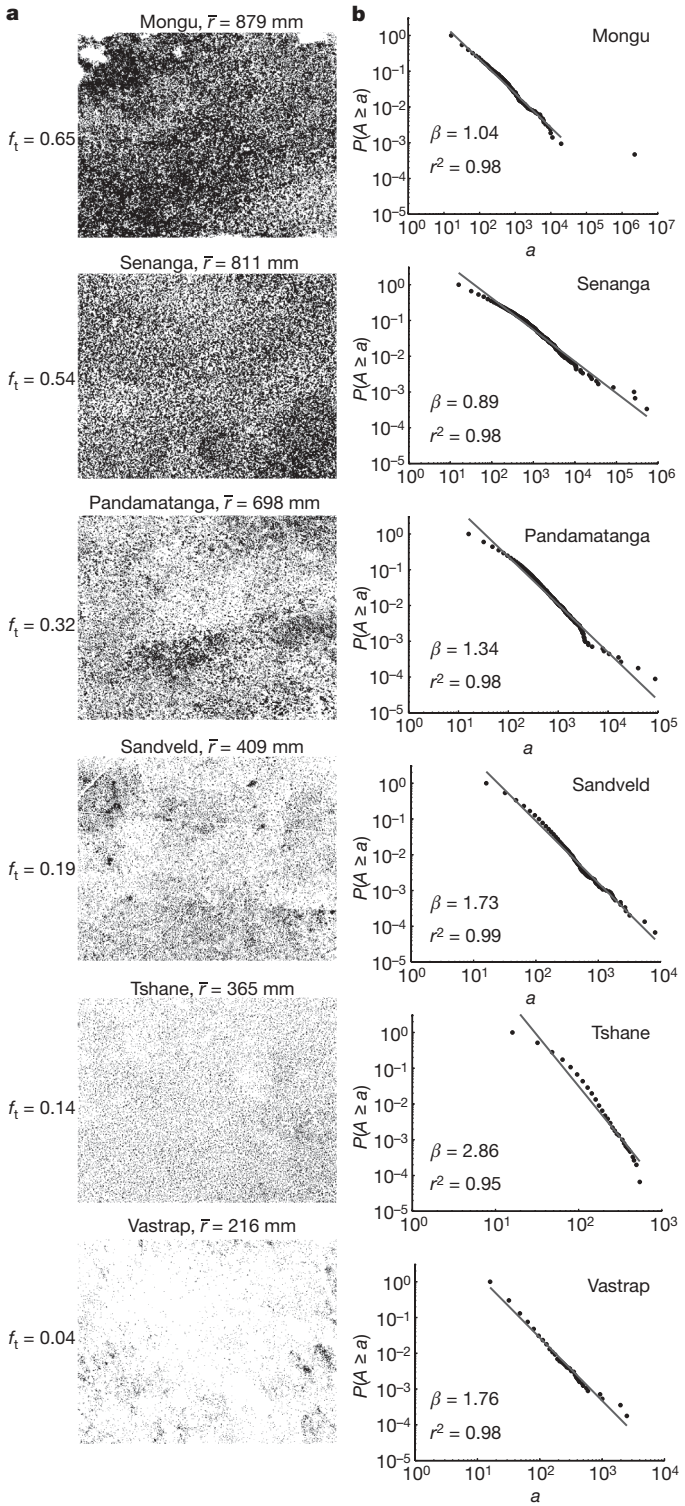


Figure 1 | Satellite observations of tree canopies and cluster size distributions. a, Binary data showing map views of the remotely sensed tree canopies, in which black points refer to the location of trees. The overall field of view is 2 km × 2 km, and the resolution is 4 m. Tree canopies were classified by thresholding the normalized difference vegetation index in the IKONOS scenes to match the fractional tree cover (f_t) measured in the field at each site. **b**, Cluster analysis of the tree canopy matrices, plotted as an inverse cumulative distribution on log–log axes. Power-law clustering is evident for a majority of the Kalahari sites, which vary widely in tree fractional cover along the rainfall gradient.

determine the size of contiguous clusters, we defined 'connected' tree pixels as those connected through any shared edge (that is, von Neumann neighbourhood; four immediate neighbours, no diagonals). In Fig. 1b, we provide the probability distribution for each of the six sites. These distributions demonstrate power-law relationships conforming to $P(A \geq a) \propto a^{-\beta}$ over a wide range of scales at all sites, with the possible exception of Tshane, where the cluster size distribution more closely resembles an exponential relationship. At the Mongu site, the single cluster that lies outside the power relationship corresponds to a cluster that spans the entire image being analysed. Power-law cluster size distributions such as these observed in the Kalahari have been cited as evidence of criticality[14], in which small perturbations to a forcing variable can lead to rapid and widespread changes to the ecosystem state (that is, savanna tree cover).

The ubiquity of power-law distributions of tree canopy cluster sizes merits further scrutiny. The coefficient of determination R^2 may be relatively high when fitting log–log relationships without any dynamic causality. In addition, percolation theory predicts the existence of power-law cluster size distributions for uniform random patterns at a fractional cover of approximately 0.59 for a square lattice with the von Neumann neighbourhood[15]. To determine the significance of our results and to provide a benchmark for our subsequent modelling analyses, we compare our observed distributions at each site to those produced by a neutral model[16]. The neutral model generates patterns by randomly assigning occupancy within a 500 × 500 matrix until the percentage of occupied pixels matches the overall site percentage cover. In this way, the neutral model matches the global constraint on tree cover, but contains no additional spatial processes. At each site we estimate the probability distribution function resulting from each of 1,000 simulations.

As shown in Table 2, the resulting random neutral models are largely unsuccessful in meeting the criteria of having both (1) an R^2 value greater than 0.98 (as exhibited by most of the satellite data), and (2) a size distribution spanning at least one and a half orders of magnitude. The latter condition was imposed to eliminate seemingly strong power-law relationships that can result from a deficiency of scales[17]. Cluster distributions derived from the satellite data demonstrate a tendency to form power-law distribution over a broad range of fractional covers, including at densities far from the percolation threshold (for more general fits to the satellite data, see Supplementary Table 3). The persistence of this scale-invariant tree pattern at sites along the Kalahari rainfall gradient begs a mechanistic understanding of the processes that lead to this emergent statistical property.

Power-law clustering has been observed in nature for a variety of phenomena, including mussel beds[18], forest gaps[19] and forest fires[20], and the pattern-formation processes for these systems have been evaluated through the implementation of lattice-based cellular automata models, in which complex system dynamics are represented by simple rules of interaction. What makes the present study distinct from these earlier findings is that the observed statistical pattern is maintained over an environmental gradient for a wide range of

Table 2 | Random neutral model versus cellular automata model.

		Random neutral model			Cellular automata model		
f_t	R^2_{obs}	$R^2 \geq 0.98$	$a_{max}/a_{min} > 10^{1.5}$	Both true	$R^2 \geq 0.98$	$a_{max}/a_{min} > 10^{1.5}$	Both true
0.65	0.98	802	1,000	802	849	1,000	849
0.54	0.98	0	1,000	0	944	1,000	944
0.32	0.98	0	1,000	0	993	1,000	993
0.19	0.99	0	0	0	701	1,000	701
0.14	0.94	605	0	0	652	1,000	652
0.04	0.98	455	0	0	577	969	577

Fits indicate the ability of models to produce power-law distributions, such as those detected from the satellite data. A total of 1,000 distributions were generated for each fractional cover. The number of model realizations in which the power law R^2 fit exceeded a threshold of 0.98 ($R^2 \geq 0.98$) and/or the distributions spanned at least one and a half orders of magnitude ($a_{max}/a_{min} > 10^{1.5}$) are provided.

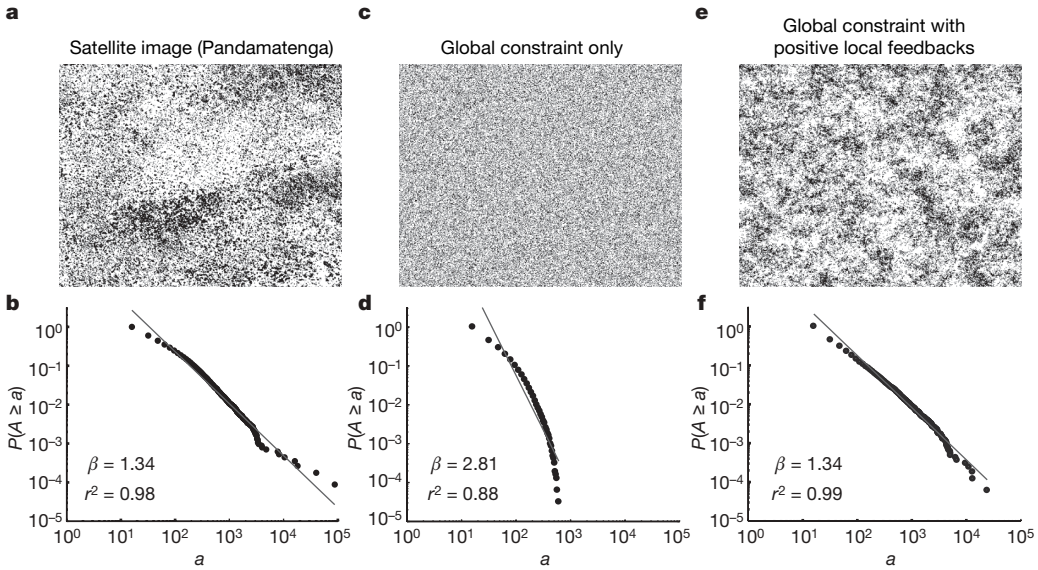


Figure 2 | Observations and models of tree canopy clustering. a, **b**, Satellite-observed tree cover for the Pandamatenga, Botswana site (**a**), and its inverse cumulative distribution that approximates a power-law fit (**b**). **c**, **d**, Spatial distribution of tree canopies produced by a random neutral model that considers only a global constraint on tree density (**c**), along with the inverse cumulative distribution (**d**), which approximates an exponential distribution of cluster sizes. **e**, **f**, Cellular automata-derived tree canopy distribution that accounts for both global constraints and positive local feedbacks on tree densities (**e**), with the scale-invariant distribution of cluster sizes resembling that of the satellite data (**f**).

vegetation densities. Any model capable of simulating the pattern-formation processes must meet the rigorous criteria of being strictly self-organizing (that is, no 'fine tuning' of parameters) and robust with respect to external environmental forcing. A recent survey of models used to produce power-law cluster size distributions in ecological systems[14] did not identify a general type that could satisfy the above conditions for a two-phase system (for example, presence/absence of trees). Disturbance-based models such as those typically used to describe power-law cluster formation do not seem to be realistic for the Kalahari setting. For instance, forest fire models[21], although capable of producing power-law cluster sizes, are not physically realistic because fire in this region is generally low-intensity and does not cause widespread mortality of trees[22].

Satellite observations and field surveys indicate that rainfall exerts a global control on tree density along the water-limited Kalahari Transect[12,23], whereas local interactions influence the spatial arrangement of individuals. A form of cellular automata model consistent with this framework is the Ising model of ferromagnetism, the two parameters of which account for global and local effects on transition dynamics. This model has previously been applied to reproduce an observed power-law gap size distribution in a neotropical forest[4]. Adapting the Ising model to account for the self-organized behaviour in the present case, however, is unsatisfactory given that it requires calibration to converge on power-law cluster size distributions. We therefore modified the model by linearizing the functional dependence of the transition probability on the neighbourhood and global structure, as well as by considering the influence of individuals beyond the von Neumann neighbourhood (see discussion in Supplementary Information). With no calibration, the model was capable of producing power-law cluster size distributions with $R^2 \geq 0.98$ at success rates of 57.7–99.3% for the six Kalahari Transect sites (Table 2). An example of the model output is shown in Fig. 2.

Power-law cluster size distributions are hallmarks of self-organization[24], and the consistent statistical pattern observed in the Kalahari points to internal feedbacks, rather than imposed spatial heterogeneity, in determining landscape-level vegetation distribution. An additional concept commonly associated with power-law cluster size distributions is criticality, which signifies a system poised at a phase transition[25]. This raises the question: is the Kalahari ecosystem in a critical state such that small perturbations could result in rapid phase change (for example, desertification) from local interactions? This is highly unlikely, because disturbance propagation is required over relatively short timescales, and there is no physically based mechanism for this in the Kalahari. Furthermore, the model presented here exemplifies power-law formation in the absence of threshold behaviour and large, temporally intermittent fluctuations. If the dynamics were to be consistent at all with a critical state it would be with the kind of 'robust' criticality recently described[14]. Climate-driven phase transitions are possible in the Kalahari, but most likely are due to the global properties of the system through positive vegetation–climate feedbacks, as reported in the Sahel region of Africa[26].

We infer that the emergent spatial pattern in the Kalahari results from positive spatial feedbacks, in which the probability of establishment increases with local tree density, and the probability of mortality increases with greater open space in the vicinity of the tree. Water availability is hypothesized to be the main driver of these positive feedbacks, as below-canopy areas remain wetter in savanna ecosystems[27] owing to reduced bare soil evaporation from shading. Direct measurements of soil moisture at a number of locations along the Kalahari Transect have confirmed this general finding. Establishment is thus favoured in areas surrounded by trees, but this positive density-dependence can also be accounted for by seed dispersal[28] and nutrient availability[29], both of which are enhanced near existing tree canopies. Mortality brought on by water stress during dry years would be more pronounced for trees that do not have the benefit of neighbourhood shading, and increased lateral hydraulic gradients would deplete the soil moisture even further for these isolated individuals. Positive feedbacks of this type could lead to either desert or fully forested conditions[30], were it not for the density-independent global effect of rainfall. This, together with the distance-weighted local effects, leads to stable power-law cluster size distributions over a wide range of vegetation densities.

METHODS SUMMARY

The cellular automata model considers both local and global effects on transition probabilities between two states: tree canopy (t) and non-tree canopy (o). The local effect is governed by the neighbourhood tree density, ρ_t, which is weighted as a function of distance, d, away from the cell undergoing possible transition

according to a Pareto-type function. The spatial 'immediacy' of the neighbourhood effect is represented by the parameter k in the Pareto-like weighting $(d_{\min}/d)^k$, where $d_{\min}$ is the minimum distance between cells in the model domain (4 m). A value of 3.0 was used for k in all simulations; this magnitude seems to affect β, but not the ability of the model to produce power-law cluster size distributions. The larger the value of k, the greater the weight placed on the tree density within the immediate vicinity of the cell in regard to its impact on ρ_t, which is defined as:

$$\rho_t = \sum_{\Omega_{i,j,M}} \left(d_{\min}/d_{i,j}\right)^k x_{i,j} \Big/ \sum_{\Omega_{i,j,M}} \left(d_{\min}/d_{i,j}\right)^k.$$

Here, $\Omega_{i,j,M}$ represents all positions i,j within a circular neighbourhood of radius M, and $x_{i,j}$ equals 1 for a tree canopy and 0 for non-tree canopy. The ability of the model to simulate power-law cluster size distributions is not contingent with the use of the Pareto weighting scheme; other functional forms representing diminished influence as a function of distance within the local neighbourhood are similarly efficient in generating power laws (see Supplementary Discussion).

The global effect on transition probability has the impact of aligning the overall fractional tree cover, f_t, with the fractional tree cover associated with the mean annual rainfall, $f_t^\star$, as determined by the observed linear relationship between these two variables. A linear combination between local and global effects yields the transition probabilities: $P(o \rightarrow t) = \rho_t + (f_t^\star - f_t)/(1 - f_t)$ and $P(t \rightarrow o) = (1 - \rho_t) + (f_t - f_t^\star)/f_t$. For each year of the simulation, 20% of the cells within the model domain were randomly selected for possible transition.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 May; accepted 2 July 2007.

1. Solé, R. V. & Manrubia, S. C. Are rainforests self-organized in a critical state? *J. Theor. Biol.* **173,** 31–40 (1995).
2. Wooten, J. Local interactions predict large-scale pattern in empirically derived cellular automata. *Nature* **413,** 841–844 (2001).
3. Levin, S. A. The problem of pattern and scale in ecology. *Ecology* **73,** 1943–1967 (1992).
4. Katori, M., Kizaki, S., Terui, Y. & Kubo, T. Forest dynamics with canopy gap expansion and stochastic Ising model. *Fractals* **6,** 81–86 (1998).
5. Condit, R. *et al.* Spatial patterns in the distribution of tropical tree species. *Science* **288,** 1414–1418 (2000).
6. Asner, G. P. & Warner, A. S. Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Rem. Sens. Environ.* **87,** 521–533 (2003).
7. Cale, W. G., Henebry, G. M. & Yeakley, J. A. Inferring process from pattern in natural communities. *Bioscience* **39,** 600–605 (1989).
8. Aguiar, M. R. & Sala, O. E. Competition, facilitation, seed distribution and the origin of patches in a Patagonian steppe. *Oikos* **70,** 26–34 (1994).
9. Archer, S. Tree-grass dynamics in a Prosopis-thornscrub savanna parkland: reconstructing the past and predicting the future. *Ecoscience* **2,** 83–99 (1995).
10. Plotkin, J. B., Chave, J. & Ashton, P. S. Cluster analysis of spatial patterns in Malaysian tree species. *Am. Nat.* **160,** 629–644 (2002).
11. Koch, G. W., Scholes, R. J., Vitousek, P. M. & Walker, B. H. *The IGBP terrestrial transects: Science plan, Report No. 36* (International Geosphere-Biosphere Programme, Stockholm, 1995).
12. Scholes, R. J. *et al.* Trends in savanna structure and composition along an aridity gradient in the Kalahari. *J. Veg. Sci.* **13,** 419–428 (2002).
13. Caylor, K. K., Shugart, H. H., Dowty, P. R. & Smith, T. M. Tree spacing along the Kalahari Transect in southern Africa. *J. Arid Environ.* **54,** 281–296 (2003).
14. Pascual, M. & Guichard, F. Criticality and disturbance in spatial ecological systems. *Trends Ecol. Evol.* **20,** 88–95 (2005).
15. Stauffer, D. & Aharony, A. *Introduction to percolation theory* (Taylor and Francis, London, 1985).
16. Keitt, T. H. Spectral representation of neutral landscapes. *Landscape Ecol.* **15,** 479–493 (2000).
17. Halley, J. M. *et al.* Uses and abuses of fractal methodology in ecology. *Ecol. Lett.* **7,** 254–271 (2004).
18. Guichard, F., Halpin, P. M., Allison, G. W., Lubchenco, J. & Menge, B. A. Mussel disturbance dynamics: Signatures of oceanographic forcing from local interactions. *Am. Nat.* **161,** 889–904 (2003).
19. Manrubia, S. C. & Solé, R. V. On forest spatial dynamics with gap formation. *J. Theor. Biol.* **187,** 159–164 (1997).
20. Malamud, B. D., Morein, G. & Turcotte, D. L. Forest fires: An example of self-organized critical behavior. *Science* **281,** 1840–1842 (1998).
21. Grassberger, P. On a self-organized critical forest-fire model. *J. Phys. A* **26,** 2081–2089 (1993).
22. Holdo, R. M. Stem mortality following fire in Kalahari sand vegetation: effects of frost, prior damage, and tree neighbourhoods. *Plant Ecol.* **180,** 77–86 (2005).
23. Scanlon, T. M., Albertson, J. D., Caylor, K. K. & Williams, C. A. Determining land surface fractional cover from NDVI and rainfall time series for a savanna ecosystem. *Rem. Sens. Environ.* **82,** 376–388 (2002).
24. Pascual, M., Roy, M., Guichard, F. & Flierl, G. Cluster size distributions: signatures of self-organization in spatial ecologies. *Phil. Trans. R. Soc. Lond. B* **357,** 657–666 (2002).
25. Yeomans, J. M. *Statistical mechanics of phase transitions* (Clarendon Press, Oxford, 1992).
26. Wang, G. L. & Eltahir, E. A. B. Biosphere–atmosphere interactions over West Africa. II: Multiple climate equilibria. *Q. J. R. Metereol. Soc.* **126,** 1261–1280 (2000).
27. Scholes, R. J. & Archer, S. R. Tree–grass interactions in savannas. *Annu. Rev. Ecol. Syst.* **28,** 517–544 (1997).
28. Nathan, R. *et al.* Mechanisms of long-distance dispersal of seeds by wind. *Nature* **418,** 409–413 (2002).
29. Schlesinger, W. H., Reikes, J. A., Hartley, A. E. & Cross, A. F. On the spatial pattern of soil nutrients in desert ecosystems. *Ecology* **77,** 364–374 (1996).
30. Molofsky, J., Bever, J. D. & Antonovics, J. Coexistence under positive frequency dependence. *Proc. R. Soc. Lond. B* **268,** 273–277 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Funding for this research was provided by grants to Princeton University from the NSF, the Mellon Foundation and the NSF National Center for Earth Surface Dynamics, and a grant to the University of Virginia from NASA IDS.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.M.S. (tms2v@virginia.edu).

METHODS

Satellite data analysis. Several of the original 9 km × 9 km IKONOS scenes of the study sites contained cloud cover and noticeable human alterations to the landscape. We avoided these effects by limiting the vegetation pattern analysis to 2 km × 2 km subsampled areas that were representative of the surrounding landscape. Red and near-infrared channels of the IKONOS images, which have a resolution of 4 m, were used to construct 500 × 500 matrices of normalized difference vegetation index (NDVI). Field measurements of fractional tree cover[12] were used to threshold the NDVI, resulting in binary matrices of the tree cover (Fig. 1a). This methodology relies on the assumption that pixels with the highest NDVI correspond to tree canopies, and the thresholding procedure effectively filters out the between-canopy areas of grass and bare soil.

Cellular automata model implementation. All model runs were initiated with 50% fractional tree cover, randomly distributed throughout a 500 × 500 model domain. Ten model runs were performed for each of the six locations along the Kalahari Transect, characterized by their respective mean annual rainfall. After spin-up periods of 200 yr, cluster size distributions were evaluated from 'snapshots' of the model output each year over 100-yr timeframes (note that 'years' represents the model time step, which should not necessarily be equated with actual time evolution).

In the numerical implementation of the Pareto weighting scheme for the local density, a value of M was chosen such that the cumulative distribution function at this radius exceeded 0.999. A linear combination of local and global effects yields the transition probabilities. In rare cases when the calculated transition probability falls outside the range {0,1}, the probability is made equal to either 0 or 1.

LETTERS

Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems

Sonia Kéfi[1], Max Rietkerk[1], Concepción L. Alados[2], Yolanda Pueyo[1], Vasilios P. Papanastasis[3], Ahmed ElAich[4] & Peter C. de Ruiter[1,5]

Humans and climate affect ecosystems and their services[1], which may involve continuous and discontinuous transitions from one stable state to another[2]. Discontinuous transitions are abrupt, irreversible and among the most catastrophic changes of ecosystems identified[1]. For terrestrial ecosystems, it has been hypothesized that vegetation patchiness could be used as a signature of imminent transitions[3,4]. Here, we analyse how vegetation patchiness changes in arid ecosystems with different grazing pressures, using both field data and a modelling approach. In the modelling approach, we extrapolated our analysis to even higher grazing pressures to investigate the vegetation patchiness when desertification is imminent. In three arid Mediterranean ecosystems in Spain, Greece and Morocco, we found that the patch-size distribution of the vegetation follows a power law. Using a stochastic cellular automaton model, we show that local positive interactions among plants can explain such power-law distributions. Furthermore, with increasing grazing pressure, the field data revealed consistent deviations from power laws. Increased grazing pressure leads to similar deviations in the model. When grazing was further increased in the model, we found that these deviations always and only occurred close to transition to desert, independent of the type of transition, and regardless of the vegetation cover. Therefore, we propose that patch-size distributions may be a warning signal for the onset of desertification.

It is of the utmost importance to find early warning signals of transitions that can alter ecosystems' services in fundamental ways, causing losses of ecological and economic resources[2,4]. Determining proximity to transitions is especially important for arid ecosystems, which may convert into deserts[2,4,5]. According to the Millennium Ecosystem Assessment, increasing external pressures by human activities or climate change will lead to desertification, affecting the livelihood of more than 25% of the world's population[1]. A mechanism playing a dominant role in the functioning of arid ecosystems is local facilitation among plants[6–9]. Local facilitation is the biophysical ameliorative effect of sessile organisms, such as plants, on their neighbouring environment. Such local positive interactions induce vegetation patchiness[6,7,10] and determine the response of this patchiness to environmental change[3].

We investigated how the spatial organization of vegetation is influenced by the degree of external stress by combining modelling and field data from three grazed Mediterranean arid ecosystems in Spain, Greece and Morocco. In each of these ecosystems, we collected data on three sites that differed with respect to the livestock grazing pressure (Table 1; Methods). In each of the nine (3×3) sites, we analysed the number and the sizes of the vegetation patches (see Methods), and plotted the number of patches, $N(S)$, as a function of their sizes, S. We fitted these patch-size distributions to two different models: a power law, $N(S) = CS^{-\gamma}$; and a truncated power law, $N(S) = CS^{-\gamma}e^{-\frac{S}{S_x}}$, where γ is the estimated scaling exponent of the model, S_x the patch size (in centimetres) above which $N(S)$ decreases faster than in a power law, and C is a constant[11,12]. To understand the mechanisms that may be responsible for the spatial organization of the vegetation, the observed distributions were compared with distributions generated by a stochastic cellular automaton model (see Methods).

We focused first on the spatial organization of the field sites with the lowest grazing pressure. In the three ecosystems, a power law best fitted the patch-size distribution characterized by a linear relation on a logarithmic scale (Fig. 1a, d and g). This power-law relation implied that vegetation patches were present over a wide range of size scales, with many small patches and relatively few large ones. The values of the scaling exponents γ of these power laws are similar among the three ecosystems, which is consistent with the hypothesis of a universal mechanism of Mediterranean ecosystem organization. At the

Table 1 | Characteristics of the three Mediterranean ecosystems

	Ecosystem type*	Dominant species	Climate	Effective stocking rate (animals per hectare per year)	Transect size (m)
Cabo de Gata-Nijar Natural Park (Almeria province, Spain)	Scattered matorral (scrubland)	*Stipa tenacissima, Chamaerops humilis, Periploca laevigata, Thymus hyemalis, Brachypodium retusum*	Average rainfall, 200 mm; mean annual temperature, 18 °C; altitude, 100 m; AI = 9.0†	Low, 0; middle, 0.27; high, 0.46	32 m (30 transects per field site)
Uta (Sithonia peninsula, Greece)	Dense matorral (shrubland)	*Cistus monspeliensis, Phillyrea latifolia, Pistacia lentiscus*	Average rainfall, 590 mm; mean annual temperature, 16.2 °C; altitude, 50 m; AI = 2.75†	Low, 0.3; middle, 2.6; high, 8.2	32 m (30 transects per field site)
Timahdit (Middle Atlas mountains, Morocco)	High mountain grassland	*Carex divisa, Genista pseudopilosa, Poa bulbosa, Thymus hyemalis*	Average rainfall, 800 mm; mean annual temperature, 22 °C; altitude, 1,900 m; AI = 2.75†	Low, 0.9; middle, 1.54; high, 2.49	16 m (30 transects per field site)

* See ref. 28 for more details.
† The aridity index[29] AI is defined as $AI = 100T/P$, where T is the average annual temperature in °C and P is the average annual rainfall in mm. AI = 2–3 characterizes a semi-arid area. AI > 3 characterizes an arid area.

[1]Department of Environmental Sciences, Copernicus Institute, Utrecht University, PO Box 80115, 3508 TC Utrecht, The Netherlands. [2]Pyrenean Institute of Ecology, Avda. Montañana 1005. Apdo. 202, 50192 Zaragoza, Spain. [3]Laboratory of Rangeland Ecology, Aristotle University, 54006 Thessaloniki, Greece. [4]Département des Productions Animales, Institut Agronomique et Vétérinaire Hassan II, Rabat, Morocco. [5]Soil Center, Wageningen University and Research Center, Droevendaalsesteeg 4, 6708 PB Wageningen, The Netherlands.

sites with higher grazing pressure, we found that a truncated power law best described the patch-size distribution: that is, there was a consistent deviation from a power law for all three ecosystems (Fig. 1b, c, e, f, h and i). The scaling exponents γ of these truncated power laws have smaller values than the γ estimated for the power-law relations. These deviations from power laws are due to a deficiency of large patches in areas described by truncated power laws compared to power laws.

To identify the mechanisms responsible for generating the power laws and their deviations, we constructed a spatial model (see Methods) of arid ecosystems. The model described arid ecosystems as lattice-structured habitats, in which each cell is occupied by vegetation (denoted +), unoccupied by vegetation (denoted 0), or degraded (denoted −). At each time step, the status of each cell can change with a probability per unit of time, depending on the status of the cell and its neighbours. Plants reproduce by spreading seeds throughout the lattice. A fraction of the seeds is dispersed locally, whereas the rest is dispersed globally[13]. The recruitment of a new individual has a probability of being successful only if the seeds reach a {0}-cell and depends on global competition for resources (negative density-dependence). The mortality of a {+}-cell occurs at a density-independent rate and may turn a {+}-cell into a {0}-cell. A {0}-cell may undergo further degradation, for example, by processes such as erosion and soil crust formation. This may turn a {0}-cell into a {−}-cell in a density-independent manner. Regeneration of a {−}-cell is faster when there are more {+}-cells in its neighbourhood, because of the positive effect of vegetation on its micro-environment; this is how local facilitation is modelled. We call a system with no vegetation cells a desert. Local positive interactions include local facilitation and local seed dispersal. We model grazing pressure as higher mortality of {+}-cells, which is the minimal possible way of including grazing in our model[14]. Grazing may include other effects, such as soil trampling or non-random movement of the animals that we do not take into account in our model. We present the model results for varying grazing pressure, but varying aridity has the same qualitative effect.

The model results showed that in systems with low grazing pressure strong local positive interactions (that is, strong local facilitation and a large proportion of seeds locally dispersed) led to a patch-size distribution characterized by a power law (Fig. 2a). When we decreased the strength of local positive interactions, the patch-size distribution deviated from a power law (Fig. 2b, c). In other words, strong local positive interactions are needed in our model to produce a power-law distribution at low grazing pressure. Without local positive interactions, power laws are never observed in our model. Thus, local positive interactions can explain the spatial organization of vegetation in the form of power laws in the three arid Mediterranean ecosystems with the lowest grazing pressure. We do not intend to prove here that local facilitation is the only possible mechanism generating the observed power laws. However, local facilitation is generally recognized as a dominant ecological mechanism driving the dynamics of arid ecosystems, and is more specifically known to operate in the three ecosystems where we collected the data[7,10]. So, our model results suggest that local facilitation is (at least) one of the mechanisms at the origin of the observed power laws.

When we increased grazing pressure in our model, while keeping the local positive interactions strong, the patch-size distribution deviated from a power law in a way similar to what was observed (Fig. 2d–f). In the model, the effect of increasing grazing pressure appeared to be similar to the effect of decreasing the strength of local positive interactions (Fig. 2). This can be understood intuitively

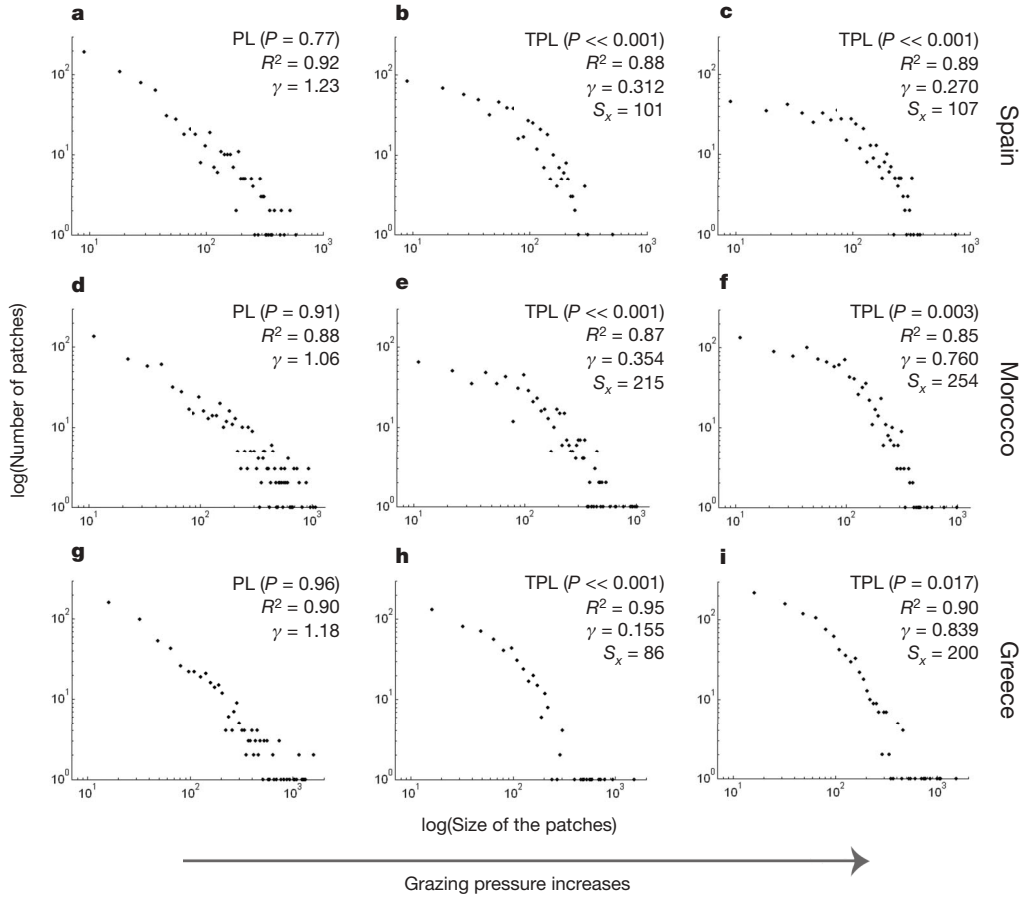


Figure 1 | Effect of grazing on the patch-size distribution of vegetation in three Mediterranean ecosystems. Increasing grazing intensity from left to right (see Table 1 for the grazing intensities). **a–c**, Spain. **d–f**, Greece. **g–i**, Morocco. The *P*-value of the sum of square reduction test, the R^2 of the best-fitted model (either power law, PL, or truncated power law, TPL), γ and S_x (see text) are given.

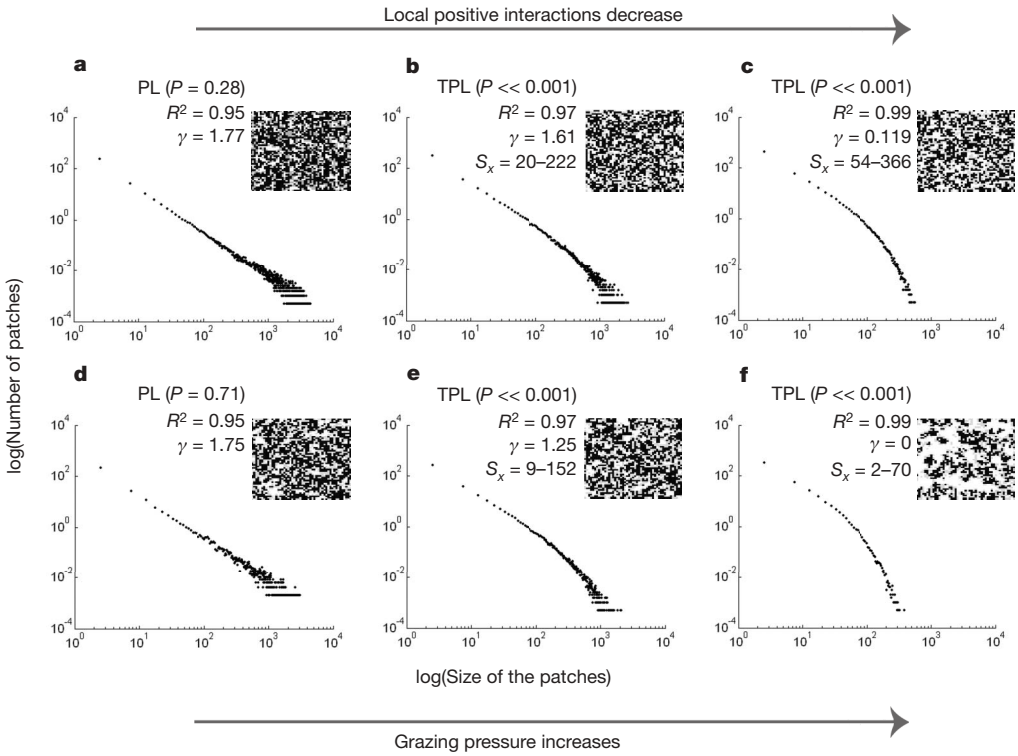


Figure 2 | Effect of local positive interactions and grazing pressure on the patch-size distribution in the model. **a–c**, Decreasing local positive interactions from left to right. **d–f**, Increasing grazing pressure from left to right. P, R^2, γ and S_x are given (see legend to Fig. 1). The first value given for S_x is a number of grid cells. Taking the square root of this value, and considering 50 cm as the length of a grid cell, leads to the second value, in centimetres. Insets are snapshots of the system at the end of the simulation. Black, vegetation; grey, recolonizable; white, degraded. See Methods for the parameter values.

because local positive interactions are crucial in maintaining vegetation in the model system, and even more so when the level of grazing pressure is high. If grazing pressure increases, stronger local positive interactions are required to maintain the system in a similar vegetation state. If the strength of local positive interactions remains the same (no plant adaptation), there is a level of grazing at which local positive interactions cannot prevent the vegetation from extinction. Thus, our model results are in agreement with our observations from Mediterranean arid ecosystems that increased grazing pressure leads to patch-size distributions that deviate from power laws, because of a decrease in the frequency of large patches.

The scaling exponents γ of the power laws obtained in the model are similar to the values observed in the data, although slightly higher (compare Fig. 2a, d with Fig. 1a, d and g). Consistent with the data, the scaling exponent γ estimated for the truncated power law is always smaller than the one estimated for the power law (Fig. 2b, c, e and f). The S_x values are of the same order of magnitude as the values estimated for the field data.

The model simulation results also showed that transitions from a vegetated to a desert state could be continuous or discontinuous (Fig. 3). At low grazing pressure, the density of vegetation gradually decreases towards zero with increasing aridity (Fig. 3a). The transition is then called continuous. However, if the grazing pressure is high, the density of vegetation undergoes discontinuous transitions with increasing aridity (Fig. 3b). This is because the system becomes bistable close to transition (Fig. 3b). Indeed, when mortality is higher

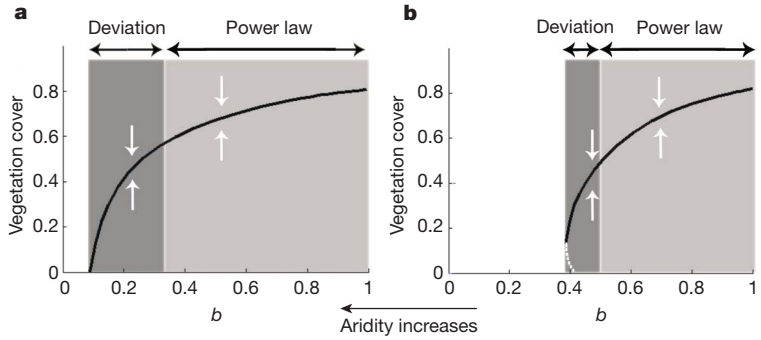


Figure 3 | Localization of the deviation to a power law along transitions in the model. Bifurcation diagrams numerically obtained from a correlation approximation of the model (Methods; see ref. 3). **a**, Continuous transition, low grazing. **b**, Discontinuous transition, high grazing. Vertical axis: fraction of the lattice occupied by vegetation at equilibrium. Horizontal axis: a lower b value reflects higher aridity. Black line, stable equilibria. Grey dashed line, unstable equilibria. Dark grey areas, deviation from power law. Light grey areas, power law. The limit value of b between the dark grey and the light grey areas was determined numerically by running the model for different b values, and by then testing which was the best-fitted model (**a**, $b = 0.34 \pm 0.01$; **b**, $b = 0.50 \pm 0.01$). White arrows indicate the direction of change of the system if the equilibrium is perturbed. See Methods for the parameter values.

because of grazing, vegetation cells might not have time to form viable patches before they die. Whether vegetation can survive in the system may then depend on the initial vegetation density, leading to bistability. In this case, a decrease in aridity to the values for which the transition occurred does not necessarily lead to a recovery of the vegetation (hysteresis). Both along continuous and discontinuous transitions, our model simulations showed that the patch-size distribution deviates from a power law just before the transition point to a desert (Fig. 3). Thus, these deviations from a power law are a general behaviour of the model system close to transition, independent of the type of transition (Fig. 3), and regardless of the vegetation cover (Fig. 2c, f).

As far as we know, power laws have not previously been described for patch-size distribution of vegetation in arid ecosystems. Power laws are commonly found in a number of biological systems[12,15–19] and physical systems (for example, see ref. 20), but the mechanisms generating them are not clearly understood in the case of ecosystems. Our study showed that local positive interactions can be responsible for power-law distributions of vegetation patches in arid Mediterranean ecosystems with a low grazing pressure. Deviations from power laws because of a deficiency of large patches in systems with high grazing pressure always occur closer to transitions than do the power laws themselves. Our model behaves differently from classical critical systems[21], where power laws only occur at the transition point. This may be explained by a major mechanistic difference between our model and classical critical models that we now address.

In classical critical systems, such as wind-disturbed tropical forests or wave-disturbed intertidal mussel beds, the abiotic disturbance (for example, wind or waves) removes the susceptible cells (occupied by trees or mussels) from the system, thereby creating disturbed cells (empty cells that are recolonizable). The intensity of the abiotic disturbance increases with the local density of disturbed cells, leading to 'active' propagation of the disturbance[22]. In those systems, the spread of the disturbance shapes the patch-size distribution. Therefore it makes sense that power laws occur for strong disturbances, corresponding to systems close to extinction. In our system, the vegetation spread (rather than the propagation of the disturbance) determines the patch-size distribution. In our system, the disturbance is a degradation process that only affects recolonizable cells creating degraded cells; so the disturbance does not directly affect the susceptible cells (in our model, the vegetation cells). The existence of a third status, recolonizable, as a necessary stage between susceptible and disturbed constrains the spatial propagation of the disturbance. We thus expect power laws to occur for high vegetation densities, corresponding to systems far from extinction.

Arid ecosystems are among the most sensitive ecosystems to global climate change[23]. High grazing pressure pushes arid ecosystems towards the edge of extinction[1]. Increased aridity can then lead to desertification in a discontinuous way, where the possibility of recovery will be low[2–4,14]. How the consistent deviations from power laws exactly relate to desertification in the field is an important and urgent question for future research, because our model results suggest that such deviations may be early warning signals for desertification of arid ecosystems.

METHODS SUMMARY

In the nine field sites, effective stocking rates (animals per hectare per year) were calculated by observation of animals. Vegetation surveys were performed on 30 random transects per site using the line-intercept method. We define a 'patch' as a distance on a transect that is totally covered by vegetation. For each site we examined the non-cumulative patch-size distribution: that is, the relationship between patch number and patch size. The two possible (nested) models for the patch-size distribution on a logarithmic scale—power law or truncated power law—were compared with a sum of square reduction test at a 5% significance level[24].

The model simulations (see Methods for model details) were carried out on grids of 100×100 cells by using a stochastic asynchronous update algorithm of the cellular automaton[25]. Two {+}-cells are part of the same patch if they have one of their four edges in common, which is consistent with the data analysis. To derive the non-cumulative patch-size distribution, we used size-classes of five cells, and evaluated the number of patches in each class. The comparison between the two models for the patch-size distributions on a logarithmic scale was done in the same way as for the data. The aridity level was estimated by b, the probability of recruitment of a new vegetation cell in a system without competition. We used the correlation approximation[26] to get numerically derived bifurcation diagrams, using the CONTENT software[3,27]. Along the transitions, we numerically determined the value of the bifurcation parameter at which the best-fitted model switches from power law to truncated power law. We did that by running the model for different aridity levels, plotting the patch-size distribution, and testing which of the two models best fits statistically.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 27 June; accepted 24 July 2007.

1. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: Desertification Synthesis* (World Resources Institute, Washington DC, 2005).
2. Scheffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413,** 591–596 (2001).
3. Kéfi, S., Rietkerk, M., van Baalen, M. & Loreau, M. Local facilitation, bistability and transitions in arid ecosystems. *Theor. Popul. Biol.* **71,** 367–379 (2007).
4. Rietkerk, M., Dekker, S. C., de Ruiter, P. C. & van de Koppel, J. Self-organized patchiness and catastrophic shifts in ecosystems. *Science* **305,** 1926–1929 (2004).
5. Reynolds, J. F. *et al.* Global desertification: Building a science for dryland development. *Science* **316,** 847–851 (2007).
6. Aguiar, M. R. & Sala, O. E. Patch structure, dynamics and implications for the functioning of arid ecosystems. *Trends Ecol. Evol.* **14,** 273–277 (1999).
7. Alados, C. L. *et al.* Association between competition and facilitation processes and vegetation spatial patterns in alpha steppes. *Biol. J. Linn. Soc.* **87,** 103–113 (2006).
8. Callaway, R. M. & Walker, L. R. Competition and facilitation: a synthetic approach to interactions in plant communities. *Ecology* **78,** 1958–1965 (1997).
9. Schlesinger, W. H. *et al.* Biological feedbacks in global desertification. *Science* **247,** 1043–1048 (1990).
10. Pugnaire, F. I., Haase, P. & Puigdefabregas, J. Facilitation between higher plant species in a semiarid environment. *Ecology* **77,** 1420–1426 (1996).
11. Jordano, P., Bascompte, J. & Olesen, J. M. Invariant properties in coevolutionary networks of plant–animal interactions. *Ecol. Lett.* **6,** 69–81 (2003).
12. Solé, R. V. & Bascompte, J. in *Self-organization in Complex Ecosystems* Ch. 6, 215–262 (Princeton Univ. Press, Princeton, 2006).
13. Iwasa, Y. in *The Geometry of Ecological Interactions. Simplify Ecological Complexity* (eds Dieckmann, U., Law, R. & Metz, J. A. J.) 227–251 (Cambridge Univ. Press, Cambridge, 2000).
14. Noy-Meir, I. Stability of grazing systems: an application of predator–prey graphs. *J. Ecol.* **63,** 459–481 (1975).
15. Brown, J. H. *et al.* The fractal nature of nature: power laws, ecological complexity and biodiversity. *Phil. Trans. R. Soc. Lond. B* **357,** 619–626 (2002).
16. Guichard, F., Halpin, P. M., Allison, G. W., Lubchenco, J. & Menge, B. A. Mussel disturbance dynamics: signatures of oceanographic forcing from local interactions. *Am. Nat.* **161,** 889–904 (2003).
17. Malamud, B. D., Morein, G. & Turcotte, D. L. Forest fires: an example of self-organized critical behavior. *Science* **281,** 1840–1842 (1998).
18. Vandermeer, J. & Perfecto, I. A keystone mutualism drives pattern in a power function. *Science* **311,** 1000–1002 (2006).
19. Venegas, J. G. *et al.* Self-organized patchiness in asthma as a prelude to catastrophic shifts. *Nature* **434,** 777–782 (2005).
20. Bak, P., Tang, C. & Wiesenfeld, K. Self-organized criticality. *Phys. Rev. A.* **38,** 364–374 (1988).
21. Sornette, D. *Critical Phenomena in Natural Sciences: Chaos, Fractals, Selforganization and Disorder: Concepts and Tools* (Springer, Berlin/Heidelberg, 2004).
22. Pascual, M. & Guichard, F. Criticality and disturbance in spatial ecological systems. *Trends Ecol. Evol.* **20,** 88–95 (2005).
23. Schröter, D. *et al.* Ecosystem service supply and vulnerability to global change in Europe. *Science* **310,** 1333–1337 (2005).
24. Schabenberger, O. & Pierce, F. J. *Contemporary Statistical Models for the Plant and Soil Sciences* Ch. 1, 1–34 (CRC Press, Boca Raton, 2002).
25. Ingerson, T. E. & Buvel, R. L. Structure in asynchronous cellular automata. *Physica D* **10,** 59–68 (1984).

26. Matsuda, H., Ogita, N., Sasaki, A. & Sato, K. Statistical mechanics of population—the lattice Lotka–Volterra model. *Prog. Theor. Phys.* **88,** 1035–1049 (1992).
27. Kutznetsov, Y. A. & Levitin, V. V. *CONTENT: a Multiplatform Environment for Continuation and Bifurcation Analysis of Dynamical Systems* (Centrum voor Wiskunde en Informatica, Amsterdam, 1997).
28. Alados, C. L. *et al.* Change in plant spatial patterns and diversity along the successional gradient of Mediterranean grazing ecosystems. *Ecol. Modell.* **180,** 523–535 (2004).
29. Pueyo, Y. & Alados, C. L. Effects of fragmentation, abiotic factors and land use on vegetation recovery in a semi-arid Mediterranean area. *Basic Appl. Ecol.* **8,** 158–170 (2007).
30. Pascual, M., Roy, M., Guichard, F. & Flierl, G. Cluster size distributions: signatures of self-organization in spatial ecologies. *Phil. Trans. R. Soc. Lond. B* **357,** 657–666 (2002).

Acknowledgements The data collection was part of the DRASME (Desertification Risk Assessment in Silvopastoral Mediterranean Ecosystems) Collaborative Research Project. DRASME is funded by the EU under the INCO-DC Program. We acknowledge the assistance of M. Vrachnakis, D. Sirkou and K. Iovi in collecting the field data in Greece. The research of S.K. and M.R. is supported by a personal VIDI grant from the Netherlands Organization of Scientific Research/Earth and Life Sciences (NWO-ALW) to M.R. The research of Y.P. is funded by the Secretaría de Estado de Universidades e Investigación of Ministerio de Educación y Ciencia (Spain). The research of P.C.d.R. is supported by the LNV-NL Strategic Research Program ''Sustainable spatial development of ecosystems, landscapes and regions''. We are grateful to M. Kéfi for his help with the figures, and to R. C. G. Chaves for commenting on the manuscript.

Author Contributions The data collection was organized and carried out by C.L.A. (Spanish, Greek and Moroccan sites), V.P.P. (Greek site) and A.E. (Moroccan site). Y.P. participated in the data collection in Spain. S.K. conducted the data analyses with help from C.L.A. and Y.P. S.K. performed the numerical simulations and analysis of the model in collaboration with M.R. and P.C.d.R., and wrote the manuscript. M.R. and P.C.d.R. supervised this work and were involved in the writing. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.K. (kefi@geo.uu.nl).

METHODS

Data collection. The data were collected in Spain, Greece and Morocco (Table 1). For each of these areas, the grazing history is well-known, and the grazing system is traditional. Three grazing pressures were identified in each area: low, medium and high. Animals (goats and sheep) were followed in the field. Their movement (located by Global Positioning System, GPS, and transferred to a map in a Geographic Information System) and the time spent in each site were recorded. Effective stocking rate was calculated as the average stocking rate multiplied by the percentage of time each grazing site was used[28]. Vegetation surveys were conducted in 2000 from April to June[28]. Transects were laid out, and the vegetation cover under the tape was recorded by measuring the starting and end points (in centimetres) of each species. We note that this is a one-dimensional measure, but the transect method is meant to give us an estimation of the patch sizes.

Data analysis. For each transect, the number and the sizes of the patches were calculated. If $P < 0.05$, the truncated power-law model was kept as the best model describing the data. For these statistical analyses, we removed the patch sizes that were only present once to minimize the effect of the transect sizes (those in Morocco were different from those in Greece and Spain). Being only present for a small fraction of the year, annuals were excluded from the analyses.

Model definition. At the cell level, our model is stochastic: each of the possible events that applies to a cell occurs with a probability per unit of time (that is, a rate). These rates can depend on the state of the four nearest neighbours. We call ρ_σ the fraction of $\{\sigma\}$-cells in the lattice (σ is $+$, 0 or $-$), and $q_{\sigma|\sigma'}$ the fraction of $\{\sigma\}$-cells in the neighbourhood of a $\{\sigma'\}$-cell. We express the colonization rate of a $\{0\}$-cell as: $w_{\{0,+\}} = \beta(\delta\rho_+ + (1-\delta)q_{+|0})G(\rho_+)$, where β represents the intrinsic seed production rate per vegetation cell multiplied by the survival and the germination probabilities; δ is the fraction of seeds dispersed all over the lattice ($1-\delta$ is the fraction dispersed in the nearest neighbourhood); G describes how seedling establishment depends on competition for resources. For simplicity, G is expressed as a linear function of ρ_+: $G(\rho_+) = \varepsilon - g\rho_+$, where ε is the establishment probability of seeds in a system without vegetation and g is the competitive effect of $\{+\}$-cells on the establishment of new individuals. Let b be $\beta\varepsilon$. We assume that mortality is density-independent: $w_{\{+,0\}} = m$. Degradation of the soil of a $\{0\}$-cell occurs at a density-independent rate: $w_{\{0,-\}} = d$. Seeds cannot germinate on $\{-\}$-cells. Regeneration of a $\{-\}$-cell occurs at rate $w_{\{-,0\}} = r + fq_{+|-}$, with r the regeneration rate of a $\{-\}$-cell without vegetation in its neighbourhood. Local facilitation f is the positive effect of neighbouring $\{+\}$-cells on regeneration. A higher local facilitation means that, for a given number of $\{+\}$-cells in the neighbourhood of a $\{-\}$-cell, the regeneration of the $\{-\}$-cell is faster.

Model analysis. Once the densities reached a stationary state, the number and size of the vegetation patches were calculated at each time step during 2,000 time steps. Patch-size distributions were obtained by averaging the values of these 2,000 time steps[30]. The size of a patch is the number of cells that constitutes it. Patch-size distributions have a different range of ordinate values in the model and in the data. In the model, we indeed averaged the distributions over 2,000 time steps, whereas in the data we only have one distribution per site. We modelled grazing pressure as a higher mortality of $\{+\}$-cells. We used the correlation approximation[26] to analyse our model[3]. This method allowed the numerical derivation of bifurcation diagrams with the software CONTENT[3,27].

Parameter values used for the simulations. In the top row of Fig. 2 (local interactions decreasing from left to right): $m = 0.15$, $b = 0.8$, $d = 0.2$, $c = 0.3$, $r = 0.0001$. In Fig. 2a, $f = 0.52$, $\delta = 0.58$. In Fig. 2b, $f = 0.45$, $\delta = 0.65$. In Fig. 2c, $f = 0.4$, $\delta = 0.7$. A qualitatively similar result is obtained if f is varied and δ remains constant. In the bottom row of Fig. 2 (grazing pressure increasing from left to right): $\delta = 0.1$, $b = 0.6$, $f = 0.9$, $d = 0.2$, $c = 0.3$, $r = 0.0001$. In Fig. 2d, $m = 0.12$. In Fig. 2e, $m = 0.13$. In Fig. 2f, $m = 0.15$. A qualitatively similar result is obtained if aridity b is varied, with increasing aridity from left to right. In Fig. 3, $r = 0.9$, $\delta = 0.1$, $c = 0.3$, $d = 0.2$, $r = 0.0001$. In Fig. 3a, $m = 0.05$. In Fig. 3b, $m = 0.1$.

LETTERS

A general integrative model for scaling plant growth, carbon flux, and functional trait spectra

Brian J. Enquist[1,2,3], Andrew J. Kerkhoff[1,4], Scott C. Stark[1], Nathan G. Swenson[1], Megan C. McCarthy[1] & Charles A. Price[1]

Linking functional traits to plant growth is critical for scaling attributes of organisms to the dynamics of ecosystems[1,2] and for understanding how selection shapes integrated botanical phenotypes[3]. However, a general mechanistic theory showing how traits specifically influence carbon and biomass flux within and across plants is needed. Building on foundational work on relative growth rate[4–6], recent work on functional trait spectra[7–9], and metabolic scaling theory[10,11], here we derive a generalized trait-based model of plant growth. In agreement with a wide variety of empirical data, our model uniquely predicts how key functional traits interact to regulate variation in relative growth rate, the allometric growth normalizations for both angiosperms and gymnosperms, and the quantitative form of several functional trait spectra relationships. The model also provides a general quantitative framework to incorporate additional leaf-level trait scaling relationships[7,8] and hence to unite functional trait spectra with theories of relative growth rate, and metabolic scaling. We apply the model to calculate carbon use efficiency. This often ignored trait, which may influence variation in relative growth rate, appears to vary directionally across geographic gradients. Together, our results show how both quantitative plant traits and the geometry of vascular transport networks can be merged into a common scaling theory. Our model provides a framework for predicting not only how traits covary within an integrated allometric phenotype but also how trait variation mechanistically influences plant growth and carbon flux within and across diverse ecosystems.

Plant functional traits are measurable morphological and physiological attributes that significantly affect whole-plant performance and thus, presumably, the main components of fitness: survival, growth and reproduction[12]. Much plant ecology and evolutionary biology aims to identify critical functional traits and to quantify their variation and covariation[13]. Such work has identified functional trait spectra (FTS)[5,7,8,14], which are correlations that describe how several functional traits—including leaf measures[8], patterns of whole-plant allocation[15], and attributes of stem hydraulics[16,17]—interrelate with each other. It is presumed that FTS[13] provide a basis for linking the ways in which organisms influence large-scale patterns in ecosystem function[1,2] and for understanding how so many species can coexist in communities[12,13]. However, theory deriving the origin of FTS and for linking FTS with whole-plant fitness, growth and ecosystem function is still unclear[2,18] (but see refs 4 and 19). Here, we argue that FTS must be interpreted within the context of integrated allometric phenotypes[18]. That is, variability and covariation of traits must be linked to whole-plant performance, such as growth.

Work on relative growth rate (RGR) and metabolic scaling theory (MST) has assumed that many trait correlations are ultimately governed by the isometric scaling[4,5,20,21] of whole-plant net biomass growth rate, dM/dt, and total plant photosynthetic (leaf) biomass, M_L (see also ref. 7). Specifically, as given by MST:

$$\frac{dM}{dt} = \dot{M} = \beta_A M_L \quad (1)$$

where β_A is an allometric constant and is the net biomass produced per unit leaf mass (see Supplementary Information). There are several key leaf traits influencing the net production of carbon and biomass. For example, variation in RGR (measured as the rate of total biomass produced per mass of plant, $g\,g^{-1}\,t^{-1}$) has traditionally[4,5,21] been linked to three key traits: (1) the leaf net carbon assimilation rate (NAR, measured in grams of C assimilated per cm^2 of leaf per unit time t: $g\,C\,cm^{-2}\,t^{-1}$); (2) the specific leaf area (the leaf area per unit leaf mass a_L/m_L, measured in $cm^2\,g^{-1}$); and (3) the leaf weight ratio (the ratio of total leaf mass to total plant mass, M_L/M). Therefore, dividing equation (1) by total mass M is equivalent to the classical decomposition[4] (for example, $RGR = NAR \times a_L/m_L \times M_L/M$) where $\beta_A = NAR \times a_L/m_L$. However, below we show that this decomposition of RGR lacks the critical traits influencing growth and the allometric dependency of leaf mass M_L.

Building on this model, using MST, we derive a trait-explicit scaling model of allometric plant growth (see Supplementary Information). The net biomass assimilation rate can be rewritten as $NAR = c\dot{A}_L/\omega$, where $\dot{A}_L$ (grams of C per cm^2 per unit time) is the leaf area specific photosynthetic rate, c is the net proportion of fixed carbon converted into biomass[22] (the carbon use efficiency, which is dimensionless), and ω is the fraction of whole-plant mass that is carbon (see Supplementary Information). Recent studies support an approximately linear relationship between NAR and $\dot{A}_L$ (see Supplementary Information), suggesting that c/ω does not covary systematically with leaf level photosynthetic rate. Using this expression[4,5,21] for NAR, the equation for whole-plant growth becomes:

$$\dot{M} = \beta_A M_L = NAR \times a_L/m_L \times M_L = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)M_L \quad (2)$$

where a_L is individual leaf area and m_L is individual leaf mass. All of the plant traits listed in equation (2) can vary intra- and interspecifically, so it is important to note that only when in the absence of parameter covariation should these values be estimated as independent averages only (see Supplementary Information). However, covariance terms in the expectation of β_A can be included by measuring multiple traits for each individual or species simultaneously.

Next we expand equation (2) by incorporating the importance of whole-plant size and allometric biomass allocation into the equation for growth rate: equation (2). MST and empirical data show that M_L scales with whole-plant mass as: $M_L = \beta_L M^{\theta}$. According to MST, the

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85719, USA. [2]The Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA. [3]Center for Applied Biodiversity, Science Conservation International, 2011 Crystal Drive, Suite 500, Arlington, Virginia 22202, USA. [4]Departments of Biology and Mathematics, Kenyon College, Gambier, Ohio 43022, USA.

Table 1 | Predicted normalization constants for several prominent scaling relationships

Plant growth quantity	Predicted functional equation from traits	Predicted normalization value	Observed normalization value
Allometric normalization for whole-plant growth rate per unit allometric mass	$\beta_G \approx \left(\frac{a_L}{m_L}\right)\left(\frac{c}{\omega}\dot{A}_L t_s^{max}\right)\beta_L$	Angiosperm $\beta_G = 2.43\ g^{1/4}\ yr^{-1}$; 95% CI = 0.44–11.92 Gymnosperm $\beta_G = 1.35\ g^{1/4}\ yr^{-1}$; 95% CI = 0.41–4.42	Angiosperm $\beta_G = 4.44\ g^{1/4}\ yr^{-1}$; 95% CI = 1.77–11.09 Gymnosperm $\beta_G = 1.36\ g^{1/4}\ yr^{-1}$; 95% CI = 0.80–2.40
Allometric normalization for whole-plant growth rate per unit leaf mass	$\beta_A \approx \left(\frac{a_L}{m_L}\right)\left(\frac{c}{\omega}\dot{A}_L t_s^{max}\right)$	Angiosperm $\beta_A = 5.53\ g\ yr^{-1}$; 95% CI = 0.74–29.27 Gymnosperm $\beta_A = 1.53\ g\ yr^{-1}$; 95% CI = 0.35–6.37	Angiosperm $\beta_A = 2.57\ g\ yr^{-1}$; 95% CI = 1.52–4.37 Gymnosperm $\beta_A = 1.07\ g\ yr^{-1}$; 95% CI = 0.71–1.63
Normalization for inverse scaling of leaf and growth time (carbon use efficiency, c)	$c \approx \frac{\omega \dot{M}_A}{\left(\frac{a_L}{m_L}\right)\dot{A}_L t_s^{max}\beta_L M^{\theta}}$	$c = 0.427$; 95% CI = 0.377–0.477	$c = 0.44$; 95% CI = 0.41–0.47
FTS normalization for the relationship between a_L/m_L and allometric leaf mass fraction, β_L	$\tau \approx \frac{\omega \dot{M}_A}{c\dot{A}_L t_s^{max} M^{\theta}}$	All taxa average = $0.0035\ cm^2\ g^{-3/4}$; 95% CI = 0.003–0.005	RMA regression = $0.004\ cm^2\ g^{-3/4}$; 95% CI = 0.003–0.006

For the first two rows, plant mass M is normalized to $M = 1$ g. Three out of four predicted scaling constants fell within the 95% confidence intervals of empirical scaling constant estimates. The other prediction—the value of β_A for angiosperms—was quite close to the observed value, falling within a factor of two of the observed value. The last two rows show the predicted values of c and τ, the prefactor (see Supplementary Information) for the trait-scaling relationship between a_L/m_L and the allometric leaf mass fraction: $a_L/m_L = \tau(\beta_L)^{-1}$. As described in the Supplementary Information, all of the growth quantities and statistics reported here are calculated from compilations of global trait, biomass and growth data.

value of θ is itself an important functional trait that reflects the geometry of the branching architecture. Its value (see Supplementary Information) ultimately controls the scaling of the number of leaves[10,23]. For seedlings and stems with few to no branchings, $\theta \approx 1$. However, for most plants larger than seedlings, MST and empirical data show $\theta \approx 3/4$ (ref. 23). Also, based on elaborations of MST[11] (Supplementary Information) we can show that the term β_L is governed by additional functional traits and plant size. Specifically, $\beta_L = M_L M^{-\theta} = M_L(\rho V)^{-\theta} = \phi_L \rho^{-\theta}$, where V is the total volume of the branching network or plant body and ρ is the whole-plant tissue density. The allometric constant $\phi_L \equiv M_L/V^{\theta}$ measures the mass of leaves per allometric volume of the plant body. Therefore, substituting for the M_L term in equation (2) yields an expanded trait-based growth law:

$$\dot{M} = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)\beta_L M^{\theta} = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)\left(\phi_L \rho^{-\theta}\right)M^{\theta} \quad (3)$$

Taking equation (3) and dividing both sides by the total mass M restores an explicit and expanded trait-based equation for RGR ($\dot{M}/M$):

$$\frac{\dot{M}}{M} = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)\left(\phi_L \rho^{-\theta}\right)M^{\theta-1} \quad (4)$$

As supported by empirical data[22], equation (4) predicts that the RGR should decrease as a function of plant size as $M^{\theta-1}$, whenever θ is less than 1.0 (refs 11, 24).

Building from the RGR literature, equations (2) to (4) effectively merge MST and FTS into a general growth law governed by key functional traits and the geometry of vascular transport networks. Furthermore, equations (3) to (4) provide a novel derivation of the MST normalization for allometric growth:

$$\beta_G \equiv \beta_L \beta_A = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)\left(\frac{M_L}{M^{\theta}}\right) = \left(\frac{c}{\omega}\dot{A}_L\right)\left(\frac{a_L}{m_L}\right)\phi_L \rho^{-\theta} \quad (5)$$

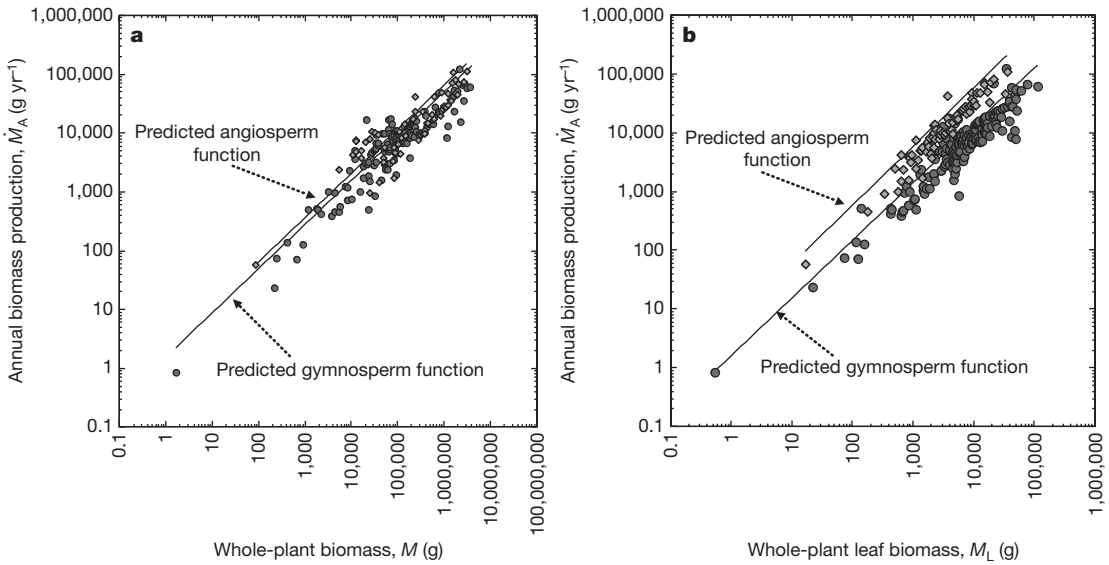


Figure 1 | Using plant traits to predict allometric growth for gymnosperms and angiosperms. a, Allometric scaling of M (roots, stems, and leaves) versus $\dot{M}_A$. **b**, Allometric scaling of M_L versus $\dot{M}_A$. Angiosperms, circles; gymnosperms, diamonds. The predicted allometric scaling functions are $\dot{M}_A = \beta_G M^{\theta}$ and $\dot{M}_A = \beta_A M_L^{1.0}$ respectively. The values of β_A and β_G were calculated for each taxon, based on resampling global values of taxon specific mean trait values as specified by equations (2) and (3) (see Supplementary Information and Table 1). Further, as discussed by MST, for trees larger than seedlings, we used the value of $\theta = 3/4$.

Table 2 | Predicted scaling exponents governing leaf trait and whole-plant allocation patterns

Functional trait scaling relationship	Predicted scaling function	Predicted exponent	Observed
a_L/m_L as a function of β_L	$\frac{a_L}{m_L} = \tau(\beta_L)^{-1}$	−1.0	−1.16; RMA 95% CI = −0.71 to −1.6 (this study; Fig. 3a)
$[(\omega\dot{M}_A)/(\beta_L M^\theta)]$ as a function of $\frac{a_L}{m_L}\dot{A}_L t_S$	$[(\omega\dot{M}_A)/(\beta_L M^\theta)] = c\left(\frac{a_L}{m_L}\dot{A}_L t_S\right)^{1.0}$	1.0	0.90; RMA 95% CI = 0.74 to 1.07 (this study; Fig. 3b)
β_L as a function of ρ	$\beta_L = \phi_L(\rho)^{-\theta}$	For terminal branches and small plants $\theta \approx -1.0$	$\theta \approx -1.22$; RMA 95% CI = −0.97 to −1.49 (Supplementary Fig. 1)
$\dot{A}_L$ as a function of ρ	$\dot{A}_L = \frac{\dot{M}\omega}{M_L c} \times \frac{m_L}{a_L}(\rho)^0$	0	0 (Supplementary Fig. 3)
a_L/m_L as a function of ρ	$\frac{a_L}{m_L} = \frac{\dot{M}\omega}{M_L c \dot{A}_L}(\rho)^0$	0	0 (data compilation from ref. 9; see also Supplementary Fig. 4)
ρ as a function of the leaf area ratio $\frac{a_L^{total}}{\phi_L M^\theta}$	$\rho = \frac{cV^\theta M_c}{\omega m_L \dot{M} t}\left(\frac{a_L^{total}}{\phi_L M^\theta}\right)^{-1/\theta}$	$-1/\theta$	Negative correlation (reported in ref. 64 in Supplementary Information)

See Supplementary Information for a detailed derivation of these and several other FTS relationships. For all of the trait correlations of *y* versus *x*, the *x*-variable trait is denoted within parentheses. Each of these trait correlations are expected to be general across all plants only if the traits in the prefactor (the traits not identified as the *x*-variable trait) do not covary with the *x*-variable trait (see section XII of the Supplementary Information). For each of these trait functional relationships: (1) both whole-plant and leaf level traits will govern the exact relationship; and (2) the geometry of the plant branching network, reflected in θ, influences each of these relationships. For each of these trait correlations, the observed empirical data generally support these predictions. The value of M_c is the mass of carbon assimilated by photosynthesis and a_L^{total} is the total leaf area of the plant (see Supplementary Information).

Thus, the value of β_G can now be predicted by measuring several key functional traits (a_L/m_L, $\dot{A}_L$, c, ω, ϕ_L, ρ and θ or M_L/M^θ, instead of $\phi_L\rho$) and, if necessary, their covariances. Additional knowledge of M would then allow us to predict growth rate.

We assessed our model in four ways. First, using equations (2) to (4) for both angiosperms and gymnosperms, we calculated the respective values of β_G and β_A by compiling global trait data sets in addition to the growing-season length, t_S^{max} (see Supplementary Information). Because growth data were on an annual basis (whole-plant biomass production $\dot{M}_A$ in units of g yr^{-1}), we approximated equation (3) by converting to an annual timescale by multiplying $\dot{A}_L$ by t_S^{max} (see Table 1 and the Supplementary Information). Confidence intervals were generated by resampling the trait distributions 100,000 times. Figure 1 and Tables 1 and 2 show that our predictions successfully approximate, with no free parameters, empirical scaling of plant growth. As supported by data, our model predicts that gymnosperms should have a higher value of β_A than angiosperms but, interestingly, both taxa have similar values of β_G due to opposing mean trait differences in a_L/m_L and $\dot{A}_L$ (Tables 1 and 2 and Fig. 1).

Second, as specified by equation (3), we assembled a smaller data set that consisted of species-specific mean trait values and individual data for plant biomass and growth. Then we used the trait data to predict the individual annual growth rates $\dot{M}_A$. Plotting the predicted $\dot{M}_A$ versus the observed $\dot{M}_A$ provides support for the model (Fig. 2).

Third, as a quantitative test, we rearranged the growth law to predict the geometric mean of a difficult-to-measure plant trait: carbon use efficiency, *c*. We calculated *c* in two ways. Rearranging equation (3) reveals that a plot of the inverse of whole-plant growth time versus the inverse of leaf physiological time (both expressed per unit leaf mass in units of t^{-1})—$[(\omega\dot{M}_A)/(\beta_L M^\theta)]$ versus $[(a_L/m_L)(\dot{A}_L t_S^{max})]$—will yield a straight line with a slope of 1.0 and an intercept the value of which is the average carbon use efficiency *c*. Empirical data generally support this prediction (Fig. 3a). The intercept of this fitted regression is the estimated value of *c* (reduced major axis regression, RMA intercept = 0.40, 95% confidence interval, CI = 0.33–0.49), which overlaps with empirical[22,25] measures of *c*. Interestingly, Fig. 3 also shows large residual variation, indicating that *c* varies between and possibly within taxa. Next, solving for *c* in equation (3), we used functional traits to calculate whole-plant *c* for several individuals. In close agreement with data[22] the value of *c* across individuals averaged 0.427 and ranged from ~0.2 to ~0.7 (see Table 1 and the Supplementary Information).

Interestingly, variation in our calculated value of *c* is positively correlated with elevation and latitude (see Supplementary Fig. 1), and negatively related to t_S^{max} ($r^2 = 0.119$, d.f. = 75, $F = 10.17$, $P = 0.002$). However, multiple regression indicates that both t_S^{max} and latitude explain 85% of variation in *c* ($r^2 = 0.726$, d.f. = 67, Akaike information criterion, AIC = −303.17). These findings are similar to those of recent studies indicating that lowland tropical forests are less carbon efficient[26] than forests having cold, short growing seasons[25,27]. Further, such variation in *c* is consistent with the hypothesis that variation in growth and tissue nutrient content across temperature gradients is adaptive[27,28], as well as that in more tropical/warm environments increased rates of herbivory and/or carbon loss to symbionts and enemies[29] may come at a cost to growth. Together, these results highlight a potential use of our model: estimating a hard-to-measure plant trait (*c*), that has profound implications for ecosystem carbon budgets, with data on growth and a handful of plant traits.

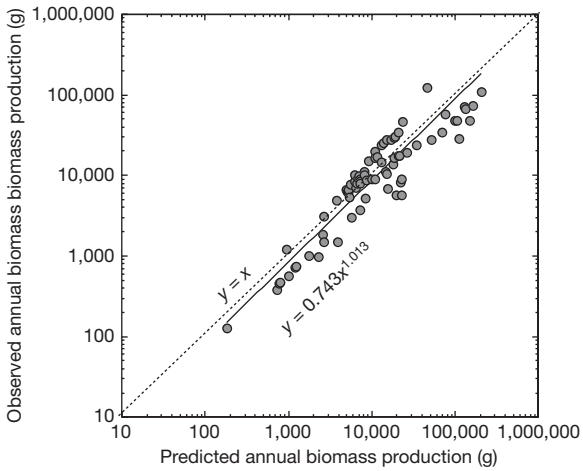


Figure 2 | Using plant traits to predict individual growth rates. For each individual, we used *M* (roots, stems, leaves) and the calculated t_S^{max} (see Supplementary Information) to calculate the predicted annual biomass production, $\dot{M}_A$. We note that the fitted RMA intercept and slope are indistinguishable from 1.0 (fitted RMA intercept = 0.743, 95% CI = 0.322–1.71; slope = 1.01, 95% CI = 0.923–1.10, $R^2 = 0.855$, $n = 79$, $P < 0.001$), indicating that predicting annual growth from trait data provides a reasonable approximation of annual growth (see Supplementary Information). The dashed line is the unity line, where the predicted values equal the observed values. Use of Ordinary Least Squares Regression does not change our results.

Fourth, we rearranged equations (3) to (4) to set one trait as a function of others to predict how traits trade off against one another and against growth rate (see Table 2 and Supplementary Information and Table 3 for a summary of key variables). Specifically, equations (3) to (4) (see also section XI of the Supplementary Information) allows us to predict, again with no free parameters, many interspecific trait correlations reported in the FTS literature including:

(1) that the specific leaf area, a_L/m_L, negatively covaries with the allometric leaf mass fraction β_L (see Fig. 4 in the Supplementary Information);

(2) that β_L scales inversely with plant tissue density with a slope of $-\theta$ (Supplementary Fig. 2);

(3); that many leaf traits including carbon flux $\dot{A}_L$ and a_L/m_L are usually independent of tissue density ρ (refs 15, 17 and Supplementary Figs 3 and 4);

(4) that ρ is negatively related to the allometric leaf area ratio (LAR, the quotient of total leaf area, $a_L^{total} = a_L N_L$, to total plant mass, where $LAR = a_L^{total}/\phi_L M^\theta$) and

(5) how these and other trait correlations (including the relationship between m_L and ρ) can also be influenced by additional trait covariation (Table 2 and Supplementary Fig. 4 and Supplementary Information).

As a result, our model provides a basis for understanding the mechanistic origin of several trait correlations and possibly trait covariation in general (see also ref. 30).

Thus, building on classic RGR studies, we have derived a general model for plant growth that integrates MST and includes many traits reported in the FTS literature. Further, the model provides a basis to understand numerous FTS trait correlations. As a result, our model shows that many of the observed correlations in the FTS literature are also influenced by additional traits detailed by MST, including plant size and the geometry of the vascular network branching architecture θ. Importantly, it also shows how allometric scaling normalizations (for example, β_L, β_A and β_G) that are not explicitly derived in MST uniquely originate in the key traits specified in FTS. In doing so, our model can then also explain the origins of vertical scatter in allometric relationships and how species differentiate from each other allometrically via differences in functional traits[12]. Furthermore, climatic or other site differences associated with functional trait variation can be linked to growth rate variation through the unified framework of our production model. Our model also indicates that studies assessing the generality of FTS must control for potential covariation in the traits listed in the scaling constants (see Supplementary Information) and that a few key plant traits (c, ϕ_L, θ and ω) have been overlooked in the FTS literature and may have important implications for plant communities and ecosystem functioning.

Our model can also be used as a point of departure for more detailed synthetic investigations (see Supplementary Information). For example, the explicit consideration of the allocation of metabolic production to reproduction and other plant organs will modify growth (see Supplementary Information). Also, the role of variation

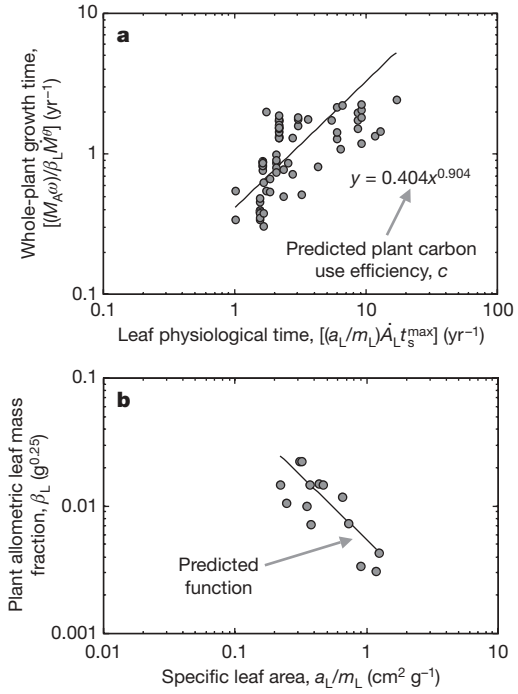


Figure 3 | Using plant traits and growth rate to predict carbon use efficiency and FTS. a, Plot of the relationship between the inverse of whole-plant growth time versus the inverse of leaf physiological time. According to our model, the best fit to the data should yield a straight line with a slope of 1.0 and an intercept the value of which is the carbon use efficiency, c (see Tables 1 and 2 and Supplementary Information). The fitted intercept, 0.40, is indistinguishable from values of c reported in the literature[22,25] (see Tables 1 and 2). We note that the large degree of residual variation indicates substantial variation in c across individuals. Indeed, much of the variation observed appears to be driven by environmental gradients (see Supplementary Fig. 1). **b**, Interspecific relationship between the leaf trait a_L/m_L and a whole-plant trait, the allometric leaf mass fraction, β_L. Our model predicts $a_L/m_L = \tau(\beta_L)^{-1}$, where $\tau \equiv \omega \dot{M}_A / c \dot{A}_L t_S^{max} M^\theta$ (measured in units of $cm^2\ g^{-3/4}$). Thus, plotting a_L/m_L versus β_L should yield a slope of –1. The model provides a testable prediction for the origin of the scaling normalization, τ. For these individuals we obtained values for the traits that govern τ. Using these traits, we predict an average value of $\tau = 0.0035$ (95% CI = −0.005 to 0.003). The predicted line, $a_L/m_L = 0.0035(\beta_L)^{-1}$, closely matches the fitted regression (see Table 1) indicating, using no free fitted parameters, that our model accurately predicts this functional relationship.

Table 3 | Definitions of key variables

Symbol	Description	Units
M	Whole-plant mass (mass of the plant's branching network including the roots, stems, and leaf petioles)	g
β_A	Allometric normalization, the net whole-plant biomass produced per unit leaf mass, $\beta_A = NAR \times a_L/m_L = \dot{M}/M_L = (c\dot{A}_L/\omega)(a_L/m_L)$	$g\,t^{-1}\,g^{-1}$
β_L	Allometric normalization, the size-weighted leaf mass fraction, where $\beta_L = M_L/M^\theta$	$g^{(1-\theta)}$
β_G	Allometric normalization for growth, where $\beta_G = \beta_L \beta_A$	$g^{(1-\theta)}$
M_L	Whole-plant leaf mass	g
$\dot{A}_L$	Area-specific carbon assimilation rate of a leaf	$g\,C\,cm^{-2}\,t^{-1}$
$\dot{m}_L$	Rate of biomass production per leaf, $\dot{m}_L = \left(\frac{c}{\omega}\dot{A}_L\right)a_L$	$g\,t^{-1}$
a_L	Area of a leaf (for example, the lamina area of a leaf)	cm^2
m_L	Mass of a leaf	g
N_L	Number of leaves per plant, $N_L = M_L/m_L$	Dimensionless
ω	Carbon fraction of plant	Dimensionless
c	Carbon use efficiency	Dimensionless
ρ	Whole plant tissue density, $\rho = M/V$	$g\,m^{-3}$
$\dot{M}$	Whole-plant biomass production rate, $\dot{M} = dM/dt = (M_L/m_L)\dot{m}_L$	$g\,t^{-1}$
$\dot{M}_A$	Whole-plant annual biomass production	$g\,yr^{-1}$
ϕ_L	Leaf mass per allometrically weighted plant volume, $\phi_L = M_L/V^\theta$	$g\,m^{-3\theta}$
t_S^{max}	Growing season length	t
θ	Composite trait calculated by the ratios of branch radii and lengths (see Supplementary Information)	Dimensionless

See also Supplementary Information for a more detailed derivation and listing.

in tissue nutrient stoichiometry (nitrogen and phosphorus), temperature and light can be incorporated via influences on $\dot{A}_L$ and c (see Supplementary Information)[7,8,28].

The development of a general quantitative theory to understand how selection can shape integrated plant phenotypes and to make more accurate predictions for the role of plants in the cycling of water, nutrients and carbon within ecosystems is central to comparative botany, physiology, and ecosystem studies. Linking RGR, MST and FTS promises to provide a general framework for explaining how the ecological and evolutionary forces that influence botanical form, function and diversity then ramify to influence the fluxes and pools of matter and energy within and across ecosystems.

Received 4 June; accepted 2 July 2007.

1. Diaz, S. & Cabido, M. Plant functional types and ecosystem function in relation to global change. *J. Veg. Sci.* **8,** 463–474 (1997).
2. Lavorel, S. & Garnier, E. Predicting changes in community composition and ecosystem functioning from plant traits: revisiting the Holy Grail. *Funct. Ecol.* **16,** 545–556 (2002).
3. Reich, P. B. *et al.* The evolution of plant functional variation: Traits, spectra, and strategies. *Int. J. Plant Sci.* **164,** S143–S164 (2003).
4. Hunt, R. *Plant Growth Analysis* (Edward Arnold Limited, London, 1978).
5. Poorter, H. in *Variation in Growth Rate and Productivity of Higher Plants* (eds Lambers, H., Cambridge, M. L., Konings, H. & Pons, T. L.) 45–68 (SPB Academic Publishing, The Hague, 1989).
6. Grime, J. P. & Hunt, R. Relatve growth-rate: its range and adaptive significance in a local flora. *J. Ecol.* **63,** 393–422 (1975).
7. Reich, P. B., Walters, M. B. & Ellsworth, D. S. From tropics to tundra: global convergence in plant functioning. *Proc. Natl Acad. Sci. USA* **94,** 13730–13734 (1997).
8. Wright, I. J. *et al.* The worldwide leaf economics spectrum. *Nature* **428,** 821–827 (2004).
9. Wright, I. J. *et al.* Relationships among major dimensions of plant trait variation in 7 neotropical forests. *Ann. Bot. Lond.* doi:10.1093/aob/mcl066 (2006).
10. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the structure and allometry of plant vascular systems. *Nature* **400,** 664–667 (1999).
11. Enquist, B. J., West, G. B., Charnov, E. L. & Brown, J. H. Allometric scaling of production and life-history variation in vascular plants. *Nature* **401,** 907–911 (1999).
12. McGill, B., Enquist, B. J., Weiher, E. & Westoby, M. Rebuilding community ecology from functional traits. *Trends Ecol. Evol.* **21,** 178–185 (2006).
13. Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A. & Wright, I. J. Plant ecological strategies: some leading dimensions of variation between species. *Annu. Rev. Ecol. Syst.* **33,** 125–159 (2002).
14. Field, C. & Mooney, H. A. in *On the Economy of Plant Form and Function* (ed. Givnish, T. J.) 25–55 (Cambridge Univ. Press, Cambridge, UK, 1986).
15. Pickup, M., Westoby, M. & Basden, A. Dry mass costs of deploying leaf area in relation to leaf size. *Funct. Ecol.* **19,** 88–97 (2005).
16. Konings, H. in *Variation in Growth Rate and Productivity of Higher Plants* (eds Lambers, H., Cambridge, M. L., Konings, H. & Pons, T. L.) 101–123 (SPB Academic Publishing, The Hague, 1989).
17. Santiago, L. S. *et al.* Leaf photosynthetic traits scale with hydraulic conductivity and wood density in Panamanian forest canopy trees. *Oecologia* **140,** 543–550 (2004).
18. Bonser, S. P. Form defining function: interpreting leaf functional variability in integrated plant phenotypes. *Oikos* **114,** 187–190 (2006).
19. Shipley, B., Lechowicz, M. J., Wright, I. & Reich, P. B. Fundamental trade-offs generating the worldwide leaf economics spectrum. *Ecology* **87,** 535–541 (2006).
20. Niklas, K. J. & Enquist, B. J. Invariant scaling relationships for interspecific plant biomass production rates and body size. *Proc. Natl Acad. Sci. USA* **98,** 2922–2927 (2001).
21. Lambers, H., Freijsen, N., Poorter, H., Hirose, T. & van der Werff, H. in *Variation in Growth Rate and Productivity of Higher Plants* (eds Lambers, H., Cambridge, M. L., Konings, H. & Pons, T. L.) 1–17 (SPB Academic Publishing, The Hague, 1989).
22. Gifford, R. M. Plant respiration in productivity models: conceptualisation, representation and issues for global terrestrial carbon-cycle research. *Funct. Plant Biol.* **30,** 171–186 (2003).
23. Enquist, B. J. *et al.* Biological scaling: Does the exception prove the rule? *Nature* **445,** E9–E10 (2007).
24. Tilman, D. *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, 1988).
25. Waring, R. H., Landsberg, J. J. & Williams, M. Net primary production of forests: a constant fraction of gross primary production? *Tree Physiol.* **18,** 129–134 (1998).
26. Chambers, J. C. *et al.* Respiration from a tropical forest ecosystem: Partitioning of sources and low carbon use efficiency. *Ecol. Appl.* **14,** S72–S88 (2004).
27. Enquist, B. J., Kerkhoff, A. J., Huxman, T. E. & Economo, E. P. Adaptive differences in plant physiology and ecosystem invariants: insights from a metabolic scaling model. *Glob. Change Biol.* **13,** 591–609 (2007).
28. Kerkhoff, A. J., Enquist, B. J., Elser, J. J. & Fagan, W. F. Plant allometry, stoichiometry and the temperature-dependence of primary productivity. *Glob. Ecol. Biogeogr.* **14,** 585–598 (2005).
29. Arnold, A. E. & Lutzoni, F. Diversity and host range of foliar endophytes: Are tropical leaves biodiversity hotspots? *Ecology* **88,** 541–549 (2007).
30. Price, C. A., Enquist, B. J. & Savage, V. M. A general model for allometric covariation in botanical form and function. *Proc. Natl Acad. Sci. USA* **104,** 13204–13209 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We especially thank J. Stegen for assistance with data and comments. We also thank S. R. Saleska, T. Huxman, A. Angert, A. E. B. Arnold, J. Pither, C. Lamanna and P. Chesson for comments and suggestions on earlier drafts. I. J. Wright provided constructive comments. B.J.E., A.J.K., C.A.P. and N.G.S. were supported by a NSF Career Award (to B.J.E.). A.J.K. was also supported by an HHMI Undergraduate Science Education Program Award to Kenyon College and N.G.S. was supported by a USGS fellowship. M.C.M. and S.C.S. were supported by an NSF predoctoral award. I. J. Wright and M. Pickup shared data sets. In addition, we acknowledge the use of GLOPNET data in some of our analyses.

Author Contributions B.J.E. designed the study. B.J.E., A.J.K. and S.C.S. developed the theory, compiled and analysed data, and wrote the paper. N.G.S., M.C.M. and C.A.P. provided data, ideas and comments on manuscript drafts.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.J.E. (benquist@email.arizona.edu).

LETTERS

Dscam diversity is essential for neuronal wiring and self-recognition

Daisuke Hattori[1]*, Ebru Demir[2]*, Ho Won Kim[1], Erika Viragh[2], S. Lawrence Zipursky[1] & Barry J. Dickson[2]

Neurons are thought to use diverse families of cell-surface molecules for cell recognition during circuit assembly. In *Drosophila*, alternative splicing of the *Down syndrome cell adhesion molecule* (*Dscam*) gene potentially generates 38,016 closely related transmembrane proteins of the immunoglobulin superfamily, each comprising one of 19,008 alternative ectodomains linked to one of two alternative transmembrane segments[1]. These ectodomains show isoform-specific homophilic binding, leading to speculation that Dscam proteins mediate cell recognition[2]. Genetic studies have established that Dscam is required for neural circuit assembly[1,3–10], but the extent to which isoform diversity contributes to this process is not known. Here we provide conclusive evidence that Dscam diversity is essential for circuit assembly. Using homologous recombination, we reduced the entire repertoire of Dscam ectodomains to just a single isoform. Neural circuits in these mutants are severely disorganized. Furthermore, we show that it is crucial for neighbouring neurons to express distinct isoforms, but that the specific identity of the isoforms expressed in an individual neuron is unimportant. We conclude that Dscam diversity provides each neuron with a unique identity by which it can distinguish its own processes from those of other neurons, and that this self-recognition is essential for wiring the *Drosophila* brain.

The complexity and specificity of neuronal wiring implies the existence of a cellular recognition code that allows neurons to distinguish between one another[11]. It has been speculated that families of highly diverse cell-surface molecules could provide this function,

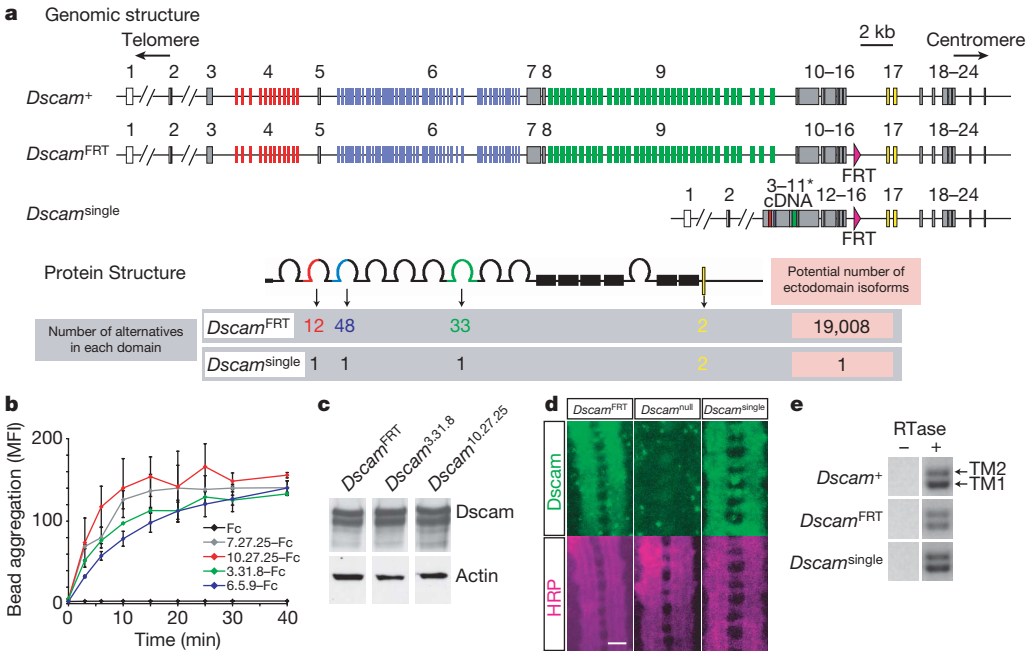


Figure 1 | Generation and molecular characterization of *Dscam*^{single} alleles. **a**, Schematic representation of the genomic organization and proteins encoded by *Dscam* alleles used in this paper. Numbers above bars in the genomic structure indicate exons. Alternatively spliced exons are shown in colour. cDNA encoding a single isoform in the *Dscam*^{single} includes exons 3–11 (asterisk). Pink triangles indicate the FRT site between exons 16 and 17. **b**, Using a bead-aggregation assay, Dscam$^{10.27.25}$, Dscam$^{3.31.8}$, and Dscam$^{6.5.9}$ ectodomains show homophilic binding similar to the Dscam$^{7.27.25}$ control[2]. Aggregation of fluorescent beads decorated with ectodomain–Fc fusion proteins was measured as an increase in the mean fluorescence intensity (MFI) of each particle. Binding experiments were performed twice. Error bars represent ± 1 s.d. **c**, *Dscam*^{single} alleles express Dscam protein at wild-type levels. The level of Dscam protein in extracts of the larval central nervous system was assessed by immunoblotting. Actin was used as a loading control. **d**, Dscam protein expression pattern (green) is normal in *Dscam*^{single} embryonic ventral nerve cord (stage 16). Anti-HRP (horseradish peroxidase) staining (purple) was used to visualize the neuropil. Scale bar, 10 μm. **e**, Expression of two alternative transmembrane (TM) domains in *Dscam*^{single} animals. RT–PCR across exons encoding each of the alternative TM domains was performed using RNA extracted from the third instar larval brains. Gel electrophoresis separates products encoding TM1 and TM2 as indicated.

[1]Department of Biological Chemistry, Howard Hughes Medical Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California 90049, USA.
[2]Institute of Molecular Pathology, Dr. Bohr-gasse 7, Vienna A-1030, Austria.
*These authors contributed equally to this work.

such as the vertebrate neurexins[12], cadherins[13] and cadherin-related neuronal receptors[14], and the insect Dscams[1,15]. However, it remains unclear to what extent the molecular diversity of such proteins is essential for wiring specificity, and how their diversity contributes to neuronal recognition. In *Drosophila melanogaster*, as many as 38,016 Dscam isoforms are generated by alternative splicing[1]. Each isoform consists of an ectodomain containing a unique combination of three different variable immunoglobulin-like domains linked to one of two alternative transmembrane segments (Fig. 1a). The variable ectodomain segments are encoded by 12, 48 and 33 alternatives for exons 4, 6 and 9, respectively, whereas the transmembrane domain is encoded by two versions of alternative exon 17 (Fig. 1a). A given ectodomain isoform binds strongly to itself, but only weakly, if at all, to other isoforms[2]. Thus, Dscam diversity could provide a molecular mechanism for selective recognition among neurons.

The potential role of Dscam diversity in neuronal wiring has previously been approached using various deletion mutations that remove subsets of alternative versions of exon 4 from the genomic locus. These alleles reduce the potential ectodomain diversity at most from 19,008 to some 5,000 isoforms. All of these mutants develop to adulthood and are fertile, and their nervous system organization is largely normal[8,16]. In one study, reducing the potential ectodomain diversity to ~11,000 isoforms resulted in an increase in the variability of axon branching and the appearance of some ectopic branches in an identified somatosensory neuron[5], alluding to a specific role for subsets of isoforms. However, because different cells splice *Dscam* differently[7,17], these genomic deletions may result in variable reduction of Dscam protein levels within different cells. For example, a specific isoform of N-cadherin is required for targeting of R7 neurons in the visual system, but this is due to cell-specific splicing rather than an isoform-specific function[18]. It is therefore unclear whether the defects observed in Dscam isoform deletion mutants reflect a functional requirement for specific isoforms, a mosaic *Dscam* loss-of-function, or both. Thus, whether Dscam diversity is essential for neural circuit assembly remains a crucial and unresolved issue. A definitive test to address the importance of ectodomain diversity would be to completely eliminate alternative splicing, reducing Dscam ectodomain diversity to just a single isoform expressed from the endogenous locus.

We used homologous recombination to replace the genomic region encoding the variable ectodomains with a complementary DNA encoding a single isoform (Fig. 1a and Supplementary Fig. 1). Three distinct ectodomains were arbitrarily selected. These were shown to exhibit homophilic binding (Fig. 1b), a property shared with all other

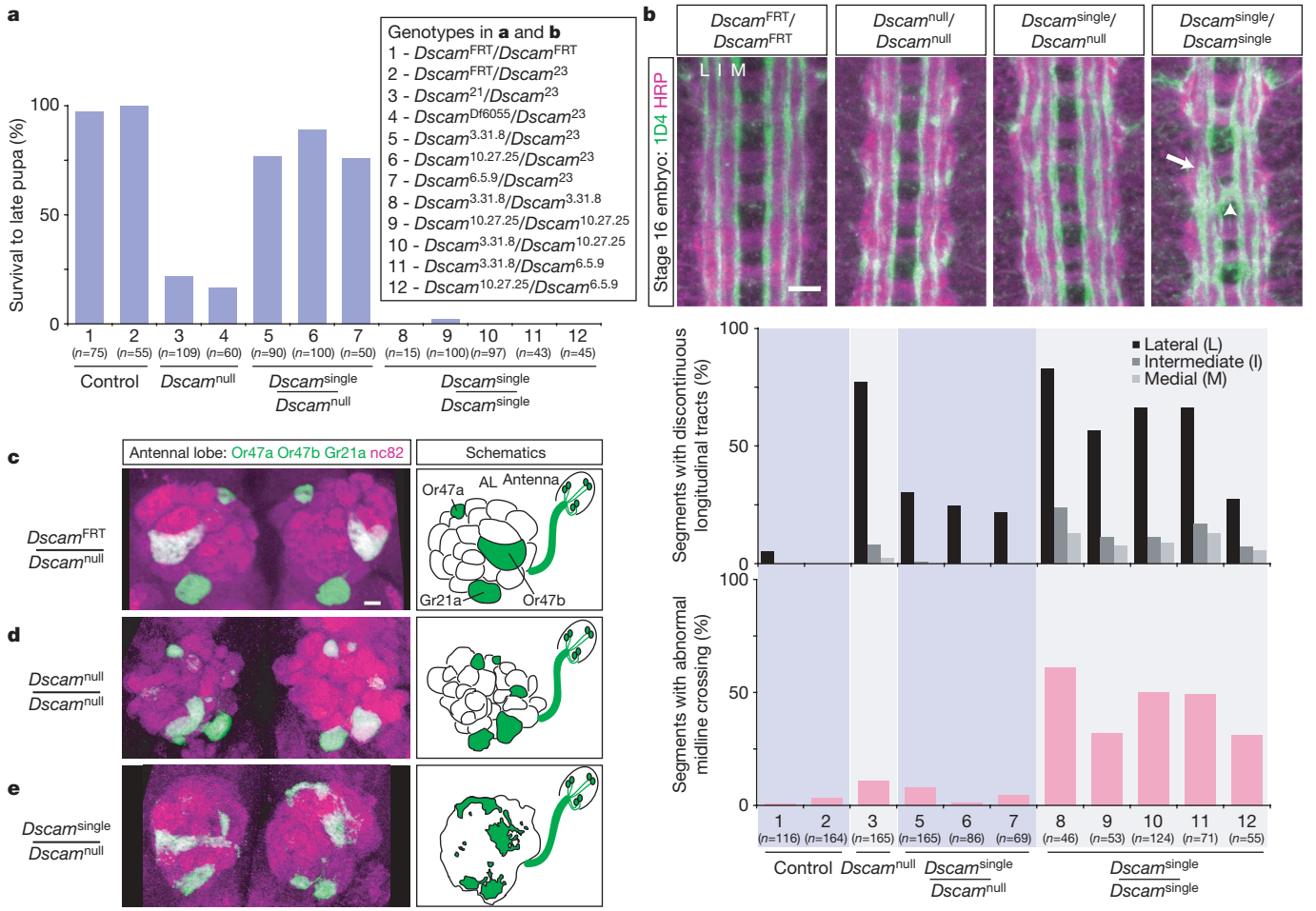


Figure 2 | Viability and neuronal wiring defects in *Dscam*^{single} mutants. **a**, Survival of different genotypes to late pupal stages. *Dscam*^{21}, *Dscam*^{23} and *Dscam*^{Df6055} are protein-null alleles. **b**, *Dscam*^{single} embryos show defects in embryonic central nervous system organization. Stage-16 embryos were examined for neuropil structure (anti-HRP, purple) and for three distinct longitudinal axon tracts (monoclonal antibody 1D4, anti FasII, green). *Dscam*^{single}/*Dscam*^{single} embryos (>95%; $n = 41$) show severely disrupted longitudinal tracts (arrow) and aberrant midline crossing (arrowhead). Scale bar, 10 μm. **c**–**e**, Dscam diversity is required in the olfactory system. Antennal lobes (AL) were visualized with the presynaptic marker nc82 (purple) and synaptotagmin–GFP (green fluorescent protein) expression (green) in three classes of olfactory receptor neurons (see schematic). **c**, *Dscam*^{FRT}/*Dscam*^{null} controls show normal organization ($n = 88$ antennal lobes). Scale bar, 10 μm. **d**, Loss of *Dscam* (*Dscam*^{null}/*Dscam*^{null}) results in mistargeting of ORNs to multiple glomeruli. The position and morphology of glomeruli are also abnormal ($n = 28$). **e**, Loss of diversity (*Dscam*^{single}/*Dscam*^{null}) leads to loss of glomerular boundaries, as well as formation of ectopic ORN termini throughout the antennal lobe. Severe phenotypes were observed for all three *Dscam*^{single} alleles (100% penetrance: $n = 28$, 128 and 16 for *Dscam*$^{3.31.8}$, *Dscam*$^{10.27.25}$ and *Dscam*$^{6.5.9}$, respectively).

isoforms we have studied (>100) (refs 2; 19). We refer to these mutant alleles collectively as *Dscam*^{single}, as the phenotypes of all three alleles were similar, and individually by indicating the selected exon variants in the allele name (for example, *Dscam*$^{3.31.8}$ contains the variable exons 4.3, 6.31, and 9.8). All of these alleles carry an FRT (FLP recombinase target) sequence inserted in the intron between exons 16 and 17. Accordingly, we also generated a control allele, *Dscam*^{FRT}, which has an FRT insertion in the same location but retains the full complement of alternative exons (Fig. 1a). We verified the intended genomic rearrangements by sequencing 14 kb from the *Dscam* locus in each of these alleles. Sequencing of cDNAs confirmed that each *Dscam*^{single} allele only expresses the designated ectodomain isoform, whereas *Dscam*^{FRT} expresses many different isoforms. Protein levels (Fig. 1c) and localization (Fig. 1d), as well as the relative use of the two alternative transmembrane exons (Fig. 1e) were similar among the *Dscam*^{single} alleles and the *Dscam*^{FRT} and wild-type controls.

Like *Dscam*^{null} alleles[1], all three *Dscam*^{single} alleles are recessive lethal. Thus, not only is Dscam itself essential, but so is its diversity.

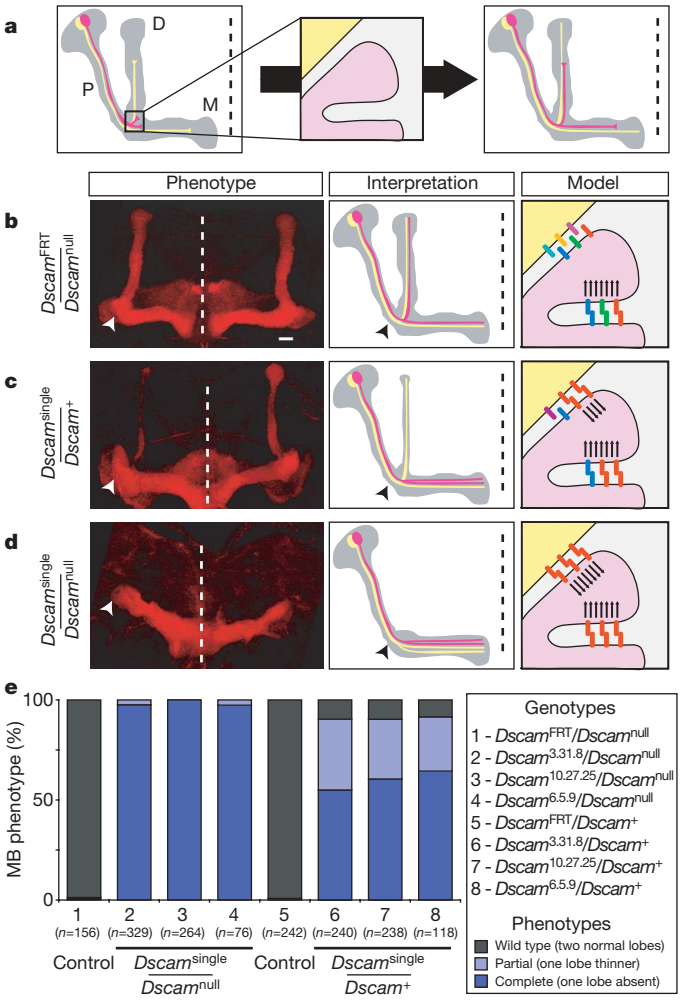


Figure 3 | *Dscam* diversity is required for mushroom body development. **a**, Schematic of mushroom body development (P, peduncle; D, dorsal lobe; M, medial lobe). Two representative mushroom body neurons are shown (yellow and pink). **b–d**, Left panels show mushroom body lobes from late pupae visualized by staining with monoclonal antibody 1D4 (anti-FasII). Only the lobe region is shown. Arrowhead indicates the location of the branch point. Dashed line, midline. Scale bar, 10 μm. Dscam isoforms expressed on sister branches (pink) and a non-self branch (yellow) are represented as coloured bars in panels on the right. Arrows indicate repulsive signals resulting from Dscam isoform-specific homophilic binding. **b**, Control animals (*Dscam*^{FRT}/*Dscam*^{null}). **c**, *Dscam*^{single}/*Dscam*$^{+}$ heterozygous animals. **d**, *Dscam*^{single}/*Dscam*^{null} mutant animals. **e**, Quantification of mushroom body (MB) phenotypes.

To assess more carefully the viability of these mutants, newly hatched mutant larvae were isolated and grown without potential competition from their siblings. Under these conditions, 17–22% of *Dscam*^{null} homozygous animals survived until late pupal stages (Fig. 2a). In contrast, as many as 76–89% of *Dscam*^{single}/*Dscam*^{null} animals survived to the pupal stages, suggesting that at least some Dscam function is diversity-independent. Notably, *Dscam*^{single} homozygotes died in early larval development, regardless of the isoform encoded (Fig. 2a). Axonal pathways were disrupted in the embryonic ventral nerve cord in all of these mutant combinations, and this phenotype too was most severe in the *Dscam*^{single} homozygotes (Fig. 2b). The phenotypes in animals carrying two different *Dscam*^{single} alleles (that is, transheterozygous animals) have a similar spectrum of phenotypes to *Dscam*^{single} homozygotes (Fig. 2a, b).

To explore further the neuronal wiring in these mutants, we examined the synaptic organization of the olfactory system (Fig. 2c–e). Olfactory receptor neurons (ORNs) project axons into the antennal lobe, where they form synapses with the dendrites of projection neurons and local interneurons. Typically, neurons in a given ORN class all express the same odorant receptor, and form connections within a distinct synaptic module, called a glomerulus[20]. Dscam is required autonomously in ORNs, projection neurons and interneurons to establish this circuitry[4,6]. In *Dscam*^{null} homozygotes, we observed a highly characteristic phenotype in which axons of a given ORN class target more than one glomerulus in each antennal lobe (Fig. 2d). Discrete glomeruli still formed in these mutants, but their stereotyped arrangement within the lobe was highly disrupted. In marked contrast, the antennal lobe of all *Dscam*^{single}/*Dscam*^{null} animals was highly disorganized with no distinct glomerular structure, and ORNs formed many ectopic termini throughout the antennal lobe (Fig. 2e). Taken together, these data establish that Dscam diversity is essential for neural circuit formation.

What is the role of Dscam diversity in circuit assembly? One possibility is that diversity is instructive, with distinct isoforms specifying distinct synaptic connections[1]. This might be an attractive model, but there is little evidence to support it. An alternative model[21] posits that Dscam diversity might provide a molecular mechanism for 'self-avoidance', that is, the propensity of multiple dendritic or axonal branches from the same neuron (sister branches) to avoid each other and thereby elaborate appropriate receptive or terminal fields, respectively[22,23]. This phenomenon requires recognition and repulsion between self-neurites, but not between neurites of different neurons. Dscam diversity might provide this function, because each neuron expresses a unique set of Dscam isoforms[7,17], and isoform-specific homophilic binding mediates repulsion[2,8]. To test this model, we focused on the simple and well-characterized sister branch segregation of mushroom body neurons in the central brain.

Each mushroom body comprises thousands of neurons. During development, each of these neurons extends an axon in a fascicle called the peduncle, at the end of which it bifurcates to produce two sister branches, one segregating into the dorsal lobe and the other into the medial lobe (Fig. 3a). Because many axons bifurcate at the same time and in close proximity to each other, each branch must be able to distinguish its sister branch from branches of other neurons, to ensure a high fidelity of sister branch segregation. Dscam is essential for this process[3], and each mushroom body neuron expresses a distinct array of Dscam isoforms[7]. To satisfy the model that Dscam diversity promotes self-recognition and self-avoidance of sister branches[21], two critical conditions must be met. First, the specific set of isoforms expressed in a given mushroom body neuron should not be important. Second, and in contrast, whatever set of isoforms a given mushroom body neuron expresses, it should be essential that this set is different from that of its neighbours. Transgenic rescue experiments and deletion mutant analysis have confirmed the first condition[7,16]. However, this condition on its own is also consistent with the null hypothesis that Dscam diversity is not required at all in mushroom body neurons. Thus, the critical test of the self-recognition model is to

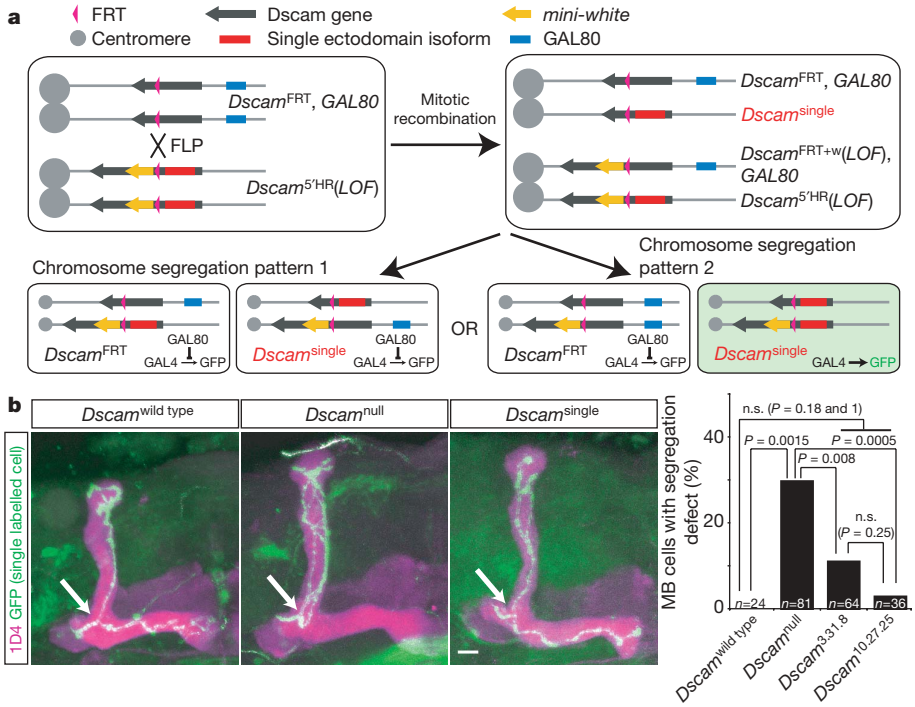


Figure 4 | *Dscam*^{single} is sufficient to promote branch segregation with high fidelity at the single-cell level. **a**, Schematic of intragenic MARCM strategy to generate and label single *Dscam*^{single} cells in an otherwise wild-type background (that is, transheterozygous with the *Dscam*^{FRT} and *Dscam* loss-of-function (*Dscam*^{LOF}) alleles). FLP recombinase induces mitotic recombination between FRT sites within the *Dscam* locus. Chromosomes can segregate in two ways. In one way (pattern 2), GFP-labelled (green) *Dscam*^{single}/*Dscam*$^{5'HR}$ mutant cells are generated. The *Dscam*$^{5'HR}$ allele (Supplementary Fig. 1) is a loss-of-function allele. If chromosomes segregate in the alternative fashion (pattern 1), no labelled cells will be produced. One of these cells will carry an intact *Dscam*^{single} allele. **b**, Branch segregation phenotypes. Labelled cells for *Dscam*$^{wild\ type}$ and *Dscam*^{single} were generated using intragenic MARCM, and *Dscam*^{null} labelled cells were produced using conventional MARCM. The mushroom body was visualized by staining with monoclonal antibody 1D4, anti-FasII (purple), and the clones were labelled with membrane-targeted chimeric GFP (mCD8GFP, green)[24]. The genotype of each clone is indicated above the panel. Arrows indicate the branch point. Quantification is shown as a bar graph (using a two-tailed Fisher's exact test; n.s., not significant).

establish whether sister branch segregation requires neighbouring mushroom body axons to express different sets of Dscam isoforms.

To test this prediction, we examined mushroom body morphology in *Dscam*^{single} and control animals (Fig. 3b–e). In *Dscam*^{FRT} controls, the normal bi-lobed structure of the mushroom body was observed, with two lobes representing populations of sister branches that had segregated correctly (Fig. 3b). In sharp contrast, this normal bi-lobed mushroom body morphology was never observed in *Dscam*^{single} animals. In some 97% of *Dscam*^{single}/*Dscam*^{null} mushroom bodies analysed, one of the two lobes was completely absent—typically the dorsal lobe (Fig. 3d). In the few remaining samples, one mushroom body lobe was significantly thinner than the other. We also observed a strong, but less severe, phenotype in ~90% of *Dscam*^{single}/*Dscam*$^{+}$ heterozygotes (Fig. 3c), with one mushroom body lobe either absent (~60%) or thinner (~30%), suggesting that some mushroom body sister branches segregate normally, but most do not. This dominant phenotype indicates that the mushroom body defects observed in *Dscam*^{single}/*Dscam*^{null} animals do not result from the loss of any one isoform, but rather the presence of the same isoform on all axons. Thus, Dscam diversity is essential for mushroom body sister branch segregation, which is compromised even if only ~50% of Dscam proteins in each neuron represent a single isoform shared by all neurons.

Although these data support a role for Dscam diversity in mushroom body branch segregation, it remains possible that the isoforms encoded in these *Dscam*^{single} alleles are simply non-functional or even inhibitory for this process. To address this issue, we generated mosaic animals in which a single mushroom body neuron expresses only one Dscam isoform from a *Dscam*^{single} allele, but all other neurons express the full complement of isoforms from the *Dscam*^{FRT} allele. This was achieved using an intragenic variation of the MARCM technique[24], in which mitotic recombination was induced between FRT sites within two modified *Dscam* alleles (Fig. 4a and Supplementary Fig. 1). We tested two different *Dscam*^{single} isoforms with this intragenic MARCM system, and found that the sister branches of isolated *Dscam*^{single} mushroom body neurons segregated with high fidelity (Fig. 4b), but that sister branches of *Dscam*^{null} neurons did not[3] (Fig. 4b). This experiment establishes that the single ectodomains we selected do indeed support sister branch segregation. Therefore, the lack of segregation in animals with a *Dscam*^{single} allele must be due to the loss of diversity, not the loss of Dscam function. Taken together, these results demonstrate that Dscam diversity is dispensable within a single neuron, but essential within a population of neurons, supporting the notion that it provides each neuron with a unique cell-surface identity.

In conclusion, here we provide strong evidence that Dscam diversity is critical for neuronal wiring. We envision that Dscam diversity contributes to wiring specificity in many different ways. One of these is self-recognition and self-avoidance, as demonstrated here for mushroom body neurons. We propose that this is a central function for Dscam diversity in neural circuit formation.

METHODS SUMMARY

The strategy used to generate the *Dscam*^{single} and *Dscam*^{FRT} alleles is indicated in Supplementary Fig. 1, and is based on the ends-in targeting strategy detailed in ref. 25. Three isoforms were selected essentially at random from sequenced cDNAs, with the only criterion being to select three distinct variants for each of exons 4, 6, and 9. Biochemical and molecular characterization of these alleles were peformed as described previously[2,7,8]. For survival analysis, each *Dscam* mutant animal was genotyped at first to second instar larval stages using *kruppel–GFP*, *CyO* balancer, and then transferred to fresh grape plates with a thin spread of yeast paste and raised at room temperature (~22–25 °C). Independently isolated alleles harbouring the same isoform were used to

generate homozygous *Dscam*^{single} animals to avoid effects from potential second site mutations. Embryonic ventral nerve cords were immunostained as described previously[7]. For other analyses, animals that survived until late pupal stage were used. Immunostainings of pupal or adult brains were performed as described previously[6,7]. For intragenic MARCM analysis, clones were generated by inducing heat-shock-mediated expression of FLP recombinase at late larval to early pupal stages. Heat-shock was carried out at 37 °C for 1 h. *Dscam*^{null} mutant clones were generated using conventional MARCM as previously described[7].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 15 March; accepted 17 July 2007.

1. Schmucker, D. *et al. Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101,** 671–684 (2000).
2. Wojtowicz, W. M., Flanagan, J. J., Millard, S. S., Zipursky, S. L. & Clemens, J. C. Alternative splicing of *Drosophila* Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. *Cell* **118,** 619–633 (2004).
3. Wang, J., Zugates, C. T., Liang, I. H., Lee, C. H. & Lee, T. *Drosophila* Dscam is required for divergent segregation of sister branches and suppresses ectopic bifurcation of axons. *Neuron* **33,** 559–571 (2002).
4. Zhu, H. *et al.* Dendritic patterning by Dscam and synaptic partner matching in the *Drosophila* antennal lobe. *Nature Neurosci.* **9,** 349–355 (2006).
5. Chen, B. E. *et al.* The molecular diversity of Dscam is functionally required for neuronal wiring specificity in *Drosophila*. *Cell* **125,** 607–620 (2006).
6. Hummel, T. *et al.* Axonal targeting of olfactory receptor neurons in *Drosophila* is controlled by Dscam. *Neuron* **37,** 221–231 (2003).
7. Zhan, X. L. *et al.* Analysis of Dscam diversity in regulating axon guidance in *Drosophila* mushroom bodies. *Neuron* **43,** 673–686 (2004).
8. Matthews, B. J. *et al.* Dendrite self-avoidance is controlled by Dscam. *Cell* **129,** 593–604 (2007).
9. Hughes, M. E. *et al.* Homophilic Dscam interactions control complex dendrite morphogenesis. *Neuron* **54,** 417–427 (2007).
10. Soba, P. *et al. Drosophila* sensory neurons require Dscam for dendritic self-avoidance and proper dendritic field organization. *Neuron* **54,** 403–416 (2007).
11. Sperry, R. W. Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Natl Acad. Sci. USA* **50,** 703–710 (1963).
12. Missler, M. & Südhof, T. C. Neurexins: three genes and 1001 products. *Trends Genet.* **14,** 20–26 (1998).
13. Takeichi, M. *et al.* Cadherins in brain patterning and neural network formation. *Cold Spring Harb. Symp. Quant. Biol.* **62,** 505–510 (1997).
14. Kohmura, N. *et al.* Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* **20,** 1137–1151 (1998).
15. Graveley, B. R. *et al.* The organization and evolution of the dipteran and hymenopteran *Down syndrome cell adhesion molecule* (*Dscam*) genes. *RNA* **10,** 1499–1506 (2004).
16. Wang, J. *et al.* Transmembrane/juxtamembrane domain-dependent Dscam distribution and function during mushroom body neuronal morphogenesis. *Neuron* **43,** 663–672 (2004).
17. Neves, G., Zucker, J., Daly, M. & Chess, A. Stochastic yet biased expression of multiple *Dscam* splice variants by individual cells. *Nature Genet.* **36,** 240–246 (2004).
18. Nern, A. *et al.* An isoform-specific allele of *Drosophila N-cadherin* disrupts a late step of R7 targeting. *Proc. Natl Acad. Sci. USA* **102,** 12944–12949 (2005).
19. Wojtowicz, W. M. *et al.* A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. *Cell*. (in the press).
20. Jefferis, G. S. & Hummel, T. Wiring specificity in the olfactory system. *Semin. Cell Dev. Biol.* **17,** 50–65 (2006).
21. Zipursky, S. L., Wojtowicz, W. M. & Hattori, D. Got diversity? Wiring the fly brain with Dscam. *Trends Biochem. Sci.* **31,** 581–588 (2006).
22. Kramer, A. P. & Kuwada, J. Y. Formation of the receptive fields of leech mechanosensory neurons during embryonic development. *J. Neurosci.* **3,** 2474–2486 (1983).
23. Grueber, W. B., Ye, B., Moore, A. W., Jan, L. Y. & Jan, Y. N. Dendrites of distinct classes of *Drosophila* sensory neurons show different capacities for homotypic repulsion. *Curr. Biol.* **13,** 618–626 (2003).
24. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22,** 451–461 (1999).
25. Rong, Y. S. & Golic, K. G. Gene targeting by homologous recombination in *Drosophila*. *Science* **288,** 2013–2018 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Zipursky and Dickson laboratories for critical comments on the manuscript. This work was supported by grants from the Austrian Science Fund (B.J.D) and NIH (S.L.Z.). Work at the Institute of Molecular Pathology is also supported by funds from Boehringer Ingelheim GmbH. S.L.Z. is an Investigator of the Howard Hughes Medical Institute.

Author Contributions B.J.D. designed the targeting strategy, and E.D., E.V. and B.J.D. generated the *Dscam*^{single} alleles, which were verified by E.D., D.H. and H.W.K. All biochemical and phenotypic analyses were performed by D.H., together with H.W.K. The intragenic MARCM strategy was conceived by D.H., E.D., H.W.K. and S.L.Z. B.J.D., D.H. and S.L.Z. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.L.Z. (lzipursky@mednet.ucla.edu) or B.J.D. (dickson@imp.ac.at).

METHODS

Generation of *Dscam*^{single} alleles. Homology regions were amplified by PCR from genomic DNA extracted from the reference strain for the *Drosophila* genome project[26], and cloned into custom-built targeting vectors. For the single-ectodomain constructs, RT–PCR was performed on total RNA extracted from the heads of Canton-S adults, amplifying exons 3–11 and inserting this fragment in-frame into the flanking genomic region within the targeting vector, using an endogenous Asp 718 site in exon 3 and an *Nhe*I site engineered into exon 11, without altering the predicted amino acid sequence. Donor insertions on the X or 3rd chromosome were obtained by P-element-mediated transformation. These donor elements were excised and linearized with *hsFLP* and *hs*I-*Sce*I, respectively[25]. The resulting virgin females were crossed to *eyFLP* males[27], and reintegration of the targeting element was detected in the progeny by the presence of an eyFLP-resistant *white*$^{+}$ marker (that is, flanked by a single FRT, rather than the two FRTs of the donor element). The final alleles were then obtained by FLP-induced recombination, as indicated in Supplementary Fig. 1, selecting for loss of the *white*$^{+}$ marker. Two independently derived isolates were established for *Dscam*$^{3.31.8}$ and *Dscam*$^{10.27.25}$ alleles, obtained in turn from two independently derived intermediate alleles. One allele was generated for *Dscam*$^{6.5.9}$.

Biochemical and molecular characterization. The bead aggregation assay was performed as previously described[2]. For immunoblots, approximately five brains dissected from the third instar larvae were homogenized directly in SDS sample buffer and western blots were performed using rabbit anti-Dscam antibody[1] at 1:1,000 dilution. The membrane was re-probed with a 1:1,000 dilution of rabbit anti-actin antibody (Sigma). RT–PCR analysis to characterize use of alternative ectodomain isoforms was performed as previously described[7]. The use of exons 17.1 and 17.2, encoding alternative transmembrane domains, was characterized by RT–PCR with primers against exon 16 and exon 18, followed by gel electrophoresis separation of PCR products containing exon 17.1 or 17.2.

Immunohistology. The following antibodies were used for immunohistology: rabbit anti-Dscam (1:500; ref. 7), monoclonal antibody 1D4 (anti-FasII; 1:10), rabbit anti-GFP (1:1,000, Molecular Probes), monoclonal antibody nc82 (1:10; ref. 28), Cy5-conjugated goat anti-HRP (1:200, Jackson ImmunoResearch Laboratories) and Alexa488- or Alexa568-conjugated goat anti-mouse or anti-rabbit (1:200, Molecular Probes). Late pupal or adult brains were dissected, and immunostaining was performed essentially as previously described[6,7]. Stage-16 embryos were fixed and immunostained as previously described[7]. Immunofluorescent samples were analysed using an LSM 510 Meta (Zeiss).

Fly stocks. *Dscam*-null alleles (*Dscam*^{21}, *Dscam*^{23} and *Dscam*^{Df6055}) have been described previously[1,6]. For antennal lobe phenotypic analyses, transgenic lines were generated, each containing a putative promoter of an *Or* gene (for example, *Or47a*) fused to a synaptotagmin–GFP marker (for example, *Or47a–sytGFP*). Three *Or–sytGFP* transgenes were used for this study (*Or47a–sytGFP*, *Or47b–sytGFP* and *Gr21a–sytGFP*) (X. L. Zhan, P. Cayirlioglu, I. C. Grunwald, D. Gunning and S.L.Z, unpublished reagents).

26. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster. Science* **287**, 2185–2195 (2000).
27. Newsome, T. P., Asling, B. & Dickson, B. J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
28. Stortkuhl, K. F., Hofbauer, A., Keller, V., Gendre, N. & Stocker, R. F. Analysis of immunocytochemical staining patterns in the antennal system of *Drosophila melanogaster. Cell Tissue Res.* **275**, 27–38 (1994).

LETTERS

Glucose sensing by POMC neurons regulates glucose homeostasis and is impaired in obesity

Laura E. Parton[1]*, Chian Ping Ye[1]*, Roberto Coppari[1,2]*, Pablo J. Enriori[3]*, Brian Choi[1], Chen-Yu Zhang[1,4], Chun Xu[3], Claudia R. Vianna[1], Nina Balthasar[1]†, Charlotte E. Lee[1], Joel K. Elmquist[2], Michael A. Cowley[3] & Bradford B. Lowell[1]

A subset of neurons in the brain, known as 'glucose-excited' neurons, depolarize and increase their firing rate in response to increases in extracellular glucose. Similar to insulin secretion by pancreatic β-cells[1], glucose excitation of neurons is driven by ATP-mediated closure of ATP-sensitive potassium (K_{ATP}) channels[2–5]. Although β-cell-like glucose sensing in neurons is well established, its physiological relevance and contribution to disease states such as type 2 diabetes remain unknown. To address these issues, we disrupted glucose sensing in glucose-excited pro-opiomelanocortin (POMC) neurons[5] via transgenic expression of a mutant Kir6.2 subunit (encoded by the *Kcnj11* gene) that prevents ATP-mediated closure of K_{ATP} channels[6,7]. Here we show that this genetic manipulation impaired the whole-body response to a systemic glucose load, demonstrating a role for glucose sensing by POMC neurons in the overall physiological control of blood glucose. We also found that glucose sensing by POMC neurons became defective in obese mice on a high-fat diet, suggesting that loss of glucose sensing by neurons has a role in the development of type 2 diabetes. The mechanism for obesity-induced loss of glucose sensing in POMC neurons involves uncoupling protein 2 (UCP2), a mitochondrial protein that impairs glucose-stimulated ATP production[8]. UCP2 negatively regulates glucose sensing in POMC neurons. We found that genetic deletion of *Ucp2* prevents obesity-induced loss of glucose sensing, and that acute pharmacological inhibition of UCP2 reverses loss of glucose sensing. We conclude that obesity-induced, UCP2-mediated loss of glucose sensing in glucose-excited neurons might have a pathogenic role in the development of type 2 diabetes.

POMC neurons in the arcuate nucleus of the hypothalamus have recently been shown to be excited by glucose[5]. The mechanism of excitation is predicted to involve ATP-induced closure of K_{ATP} channels in the plasma membrane[2,3,5]. To test this, and to create mice with defective glucose sensing in POMC neurons, we generated transgenic mice expressing a mutant form of the K_{ATP} channel subunit Kir6.2 (Kir6.2[ΔN2–30,K185Q]–GFP) (ref. 7) under transcriptional control of the mouse POMC promoter[9] (Fig. 1a). Mutant Kir6.2 forms functional K_{ATP} channels that are 250 times less sensitive to closure by ATP[7] and, when expressed in pancreatic β-cells, causes impaired glucose-induced insulin secretion and diabetes[6]. The carboxy (C)-terminal end of the mutant Kir6.2 contains a green fluorescent protein (GFP) tag that does not alter the function of the channel but makes it possible to visualize cells expressing mutant Kir6.2 (mut-Kir6.2).

To validate that the POMC-mut-Kir6.2 mice expressed the transgene only in POMC neurons, we used polymerase chain reaction with reverse transcription (RT–PCR) to probe for Kir6.2[ΔN2–30,K185Q]–GFP mRNA. We also performed double immunofluorescence for β-endorphin (a marker for POMC neurons) and GFP in brain sections of POMC-mut-Kir6.2 mice. Kir6.2[ΔN2–30,K185Q]–GFP messenger RNA was detected in the hypothalamus and pituitary only (Supplementary Fig. S1a). The signal in the pituitary is expected, as POMC is also expressed in corticotrophs and melanotrophs. This pituitary expression seems to be without consequence, as POMC-mut-Kir6.2 mice have normal corticosterone levels (fed state, wild-type mice, $34.2 \pm 7.2\ ng\,ml^{-1}$; transgenic mice, $34.3 \pm 5.7\ ng\,ml^{-1}$; mean ± s.e.m.). In the brain, GFP was found only in POMC neurons within the arcuate nucleus. More than 95% of arcuate POMC neurons expressed GFP (Fig. 1b).

We then confirmed that expression of the mut-Kir6.2 transgene in POMC neurons resulted in expression of K_{ATP} channels that were insensitive to ATP. We performed whole-cell electrophysiological recordings on both wild-type (control; POMC-GFP) and mut-Kir6.2 POMC neurons, and measured spontaneous firing, membrane potential and K_{ATP} channel currents in response to decreases in intracellular ATP. As expected, mut-Kir6.2 POMC neurons were insensitive to changes in intracellular ATP (Supplementary Fig. S1b), and expressed functional, but ATP-insensitive, K_{ATP} channels (Supplementary Fig. S1c–e, see Supplementary Information for details). We next examined the glucose sensitivity of wild-type and mut-Kir6.2 POMC neurons. We studied wild-type POMC neurons from transgenic POMC-GFP mice in the loose-patch mode, and varied glucose levels between 3 mM and 5 mM (these glucose concentrations are consistent with those described in the hypothalamus *in vivo*[10]). In agreement with previous reports[5], we found that a subpopulation of POMC neurons in wild-type mice is excited by glucose. Approximately 50% of POMC neurons (30 out of 59) reversibly decreased their firing rate as the extracellular glucose concentration was taken from 5 to 3 mM (Fig. 1c, left panel). As a group, the glucose-excited neurons had a 2.26 ± 0.23-fold faster firing rate in 5 mM glucose compared to 3 mM glucose. The excitatory effect of glucose is likely to be direct, as glucose was still able to excite POMC neurons when synaptic transmission was blocked with a low $[Ca^{2+}]$, high $[Mg^{2+}]$ solution (data not shown). Excitation of POMC neurons by glucose was almost entirely lost in POMC-mut-Kir6.2 mice, for which 26 out of 27 neurons failed to increase their firing rate in response to 5 mM glucose (Fig. 1c, middle panel). The percentage of POMC neurons excited by glucose in wild-type and POMC-mut-Kir6.2 mice is summarized in Fig. 1c (right panel).

[1]Department of Medicine, Division of Endocrinology, Beth Israel Deaconess Medical Center and Harvard Medical School, 99 Brookline Avenue, Boston, Massachusetts 02215, USA. [2]Department of Internal Medicine, Center for Hypothalamic Research, The University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390-9077, USA. [3]Division of Neuroscience, Oregon National Primate Research Center, Oregon Health & Science University, 505 NW 185th Avenue, Beaverton, Oregon 97006, USA. [4]State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210093, China. †Present address: Department of Physiology and Pharmacology, University of Bristol, Bristol BS8 1TD, UK.
*These authors contributed equally to this work.

Whole-cell current-clamp recording of membrane potential in POMC neurons showed that at low (3 mM) glucose concentrations, POMC-mut-Kir6.2 neurons are indistinguishable from wild-type neurons in membrane potential and basal firing rates, but unlike wild-type POMC neurons, fail to depolarize and to increase firing rate under conditions of raised (5 mM) glucose (Supplementary Fig. S2a, b). Failure to depolarize and increase firing rates in 5 mM glucose is consistent with a disrupted response of K_{ATP} channels to increased ATP, as expected from expression of the Kir6.2[ΔN2–30,K185Q] mutation. Leptin depolarized and increased the firing rates of 14 out of 17 wild-type neurons and 14 out of 18 mutant neurons (Supplementary Fig. S2c). These results indicate that 'leptin sensing' by POMC-mut-Kir6.2 neurons is normal and, as expected, that the electrophysiological effects of leptin on POMC neurons do not require ATP-mediated regulation of K_{ATP}-channels. Thus, glucose sensing, but not leptin sensing, is disrupted in POMC neurons of POMC-mut-Kir6.2 mice.

To test glucose-sensing by POMC neurons in POMC-mut-Kir6.2 mice using a different approach, we measured the release of α-melanocyte stimulating hormone (αMSH) from hypothalamic slices[11] in response to increasing glucose concentrations. In slices from wild-type mice, glucose stimulated αMSH secretion in a dose-dependent manner (Fig. 1d). In contrast, in hypothalamic slices from POMC-mut-Kir6.2 mice, glucose did not result in stimulation of αMSH release (Fig. 1d). Together, these findings confirm that glucose sensing in POMC-mut-Kir6.2 neurons is disrupted, and provide direct evidence that ATP-induced closure of K_{ATP} channels is required for glucose-excitation of POMC neurons.

We next investigated the effects of defective glucose sensing in POMC neurons on whole-body glucose homeostasis. Of note, body weight is normal in POMC-mut-Kir6.2 mice (wild-type, 22.3 ± 0.3 g; mut-Kir6.2 mice, 22.1 ± 0.3 g; mean ± s.e.m.). Intraperitoneal glucose tolerance tests were performed on male POMC-mut-Kir6.2 mice and their wild-type littermates at eight weeks of age. After an exogenous load of glucose (1 g glucose per kg body weight), mice lacking glucose sensing in POMC neurons (POMC-mut-Kir6.2 mice) showed impaired glucose tolerance (Fig. 1e). This phenotype was also observed in a second line of POMC-mut-Kir6.2 mice (Supplementary Fig. S2d, Line 2). These findings demonstrate that glucose sensing in glucose-excited POMC neurons is required for the normal handling of a systemic glucose load. Abnormal glucose homeostasis in the face of normal body weight regulation raises the possibility that POMC neurons are heterogeneous with respect to function.

Glucose sensing in pancreatic β-cells is lost during the development of obesity-induced type 2 diabetes[12]. Given this, and given the role of POMC neurons in handing a systemic glucose load, we tested the hypothesis that obesity may also induce a similar impairment in glucose sensing in POMC neurons. We placed wild-type mice on a high-fat diet for 20 weeks and looked at the release of αMSH in hypothalamic slices in response to elevated glucose. As expected, glucose stimulated release of αMSH from hypothalamic slices of

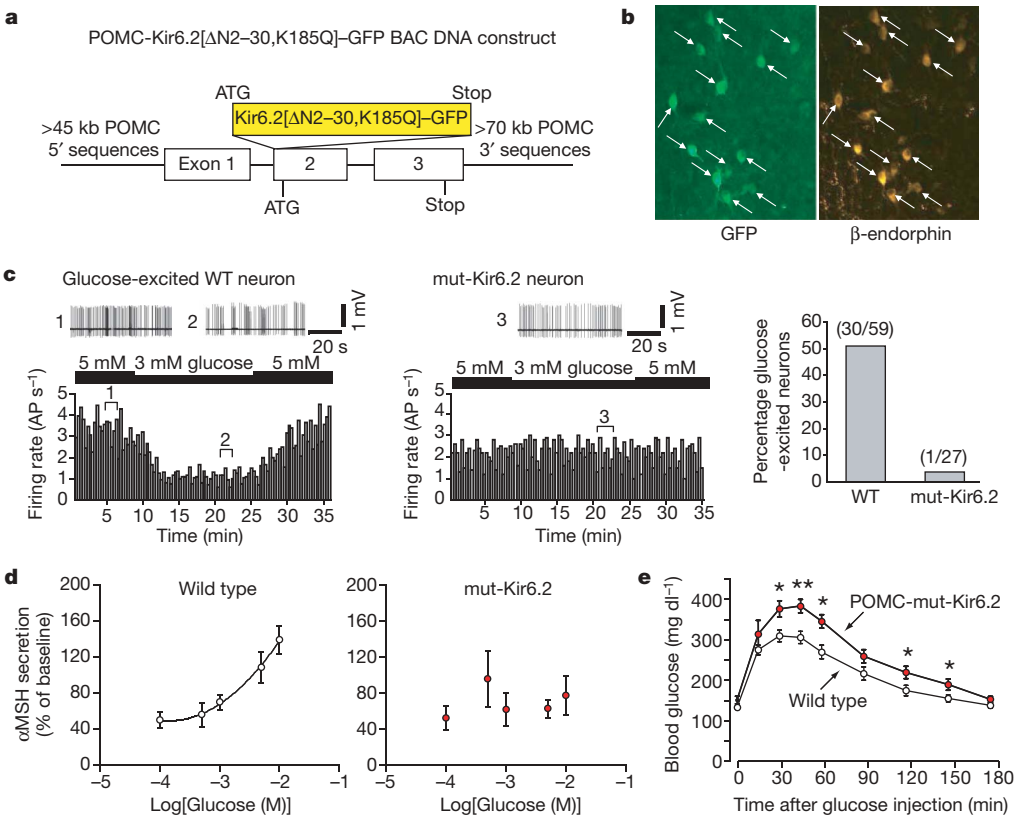


Figure 1 | Glucose sensing is lost in POMC-mut-Kir6.2 neurons. **a**, Structure of the Kir6.2[ΔN2–30,K185Q]–GFP transgene. **b**, Double immunofluorescence staining for GFP (green) and β-endorphin (yellow) in the arcuate nucleus of POMC-mut-Kir6.2 mice. Arrows indicate neurons containing both β-endorphin and Kir6.2[ΔN2–30,K185Q]–GFP. **c**, Loose-patch recordings of POMC neurons from wild-type (WT, POMC-GFP) and POMC-mut-Kir6.2 transgenic mice. Recordings were made for 5–10 min in aCSF solution containing 5 mM glucose. Once stable activities were observed, the recording chamber was perfused with aCSF solution containing 3 mM glucose for 5–15 min, then switched back to 5 mM glucose for a further 5–10 min. Panels show a representative time course of firing rate of a glucose-excited wild-type neuron (left) and a glucose-insensitive POMC-mut-Kir6.2 neuron (middle). Each bar represents the average firing rate for a 20-s interval; AP, action potential. The right panel shows the percentage of neurons activated by 5 mM glucose (recordings were obtained from 22 wild-type mice and 12 POMC-mut-Kir6.2 mice, with 2–4 POMC neurons recorded per animal). **d**, αMSH release from hypothalamic slices of wild-type and POMC-mut-Kir6.2 mice ($n = 3$ hypothalamic slices per data point, ± s.e.m.). **e**, Representative glucose tolerance curves from eight-week-old male wild-type and POMC-mut-Kir6.2 littermates ($n = 8$–10 mice per genotype, ± s.e.m.). Asterisk, $P < 0.05$; two asterisks, $P < 0.01$ compared with wild-type at a given time point.

chow-fed, wild-type mice (Fig. 2a). However, glucose failed to stimulate release of αMSH from wild-type mice fed a high-fat diet (Fig. 2a).

We also used electrophysiological techniques to independently assess glucose sensing in POMC neurons of mice on a high-fat diet. In this case, four-week-old mice were fed chow or a high-fat diet for eight weeks. Approximately 46% (5 out of 11) of POMC neurons from chow-fed, control mice significantly increased their firing rate with an increase in glucose concentration from 3 to 5 mM (fold increase, 2.17 ± 0.06), compared to only 10% of POMC neurons from mice on a high-fat diet (1 out of 10) (Fig. 2b). Notably, the fold increase in firing rate of the single glucose-responsive POMC neuron from the high-fat diet group was much lower that that of the chow-fed group (1.47-fold increase). On the basis of both the αMSH secretion assays and electrophysiological studies, we suggest that glucose sensing is defective in POMC neurons of obese mice on a high-fat diet.

In pancreatic β-cells, glucose sensing is negatively controlled by the mitochondrial protein UCP2 (ref. 13). UCP2 mediates proton leak across the inner mitochondrial membrane, decreasing the yield of ATP from glucose[8,14,15]. UCP2 activity is increased in β-cells of animal models for type 2 diabetes[8,16,17], and various studies have provided evidence that this increase in UCP2 activity has a role in the development of β-cell dysfunction[8,17–20]. Pancreatic islets from mice that lack UCP2 have higher intracellular ATP levels and are protected from chronic high glucose- and obesity-induced loss of glucose sensing[8,17,19]. UCP2 is also expressed in the brain, including in the arcuate nucleus[21–23] (Fig. 2c), and notably, in POMC neurons as well as other arcuate neurons (Fig. 2d). As glucose-excited POMC neurons sense glucose through ATP-mediated closure of K_{ATP} channels, we hypothesized that UCP2 might negatively regulate glucose sensing in POMC neurons, as it does in pancreatic β-cells. If so, UCP2 might be responsible for the loss of glucose sensing in POMC neurons observed in mice on a high-fat diet (Fig. 2a, b). Consistent with this hypothesis, *Ucp2* mRNA expression is upregulated in the hypothalami of obese mice, including those on a high-fat diet (Fig. 2e) and leptin-deficient (*ob/ob*) mice (Fig. 2f).

We have recently identified a membrane-permeant molecule, genipin, which inhibits UCP2-mediated proton leak[19]. When added to incubated pancreatic β-cells, genipin increases mitochondrial membrane potential, which then increases ATP levels and closes K_{ATP} channels, thereby stimulating insulin secretion. These actions of genipin are all secondary to inhibition of UCP2, as they are absent in β-cells from *Ucp2* knockout ($Ucp2^{-/-}$) mice[19]. Addition of genipin (20 μM) to hypothalamic slices from wild-type mice depolarized and rapidly increased the firing rate of approximately 50% of POMC neurons (15 out of 32) (Fig. 3a and Supplementary Fig. S2e). The percentage of POMC neurons excited by genipin is similar to the percentage excited by glucose. To test whether POMC neurons

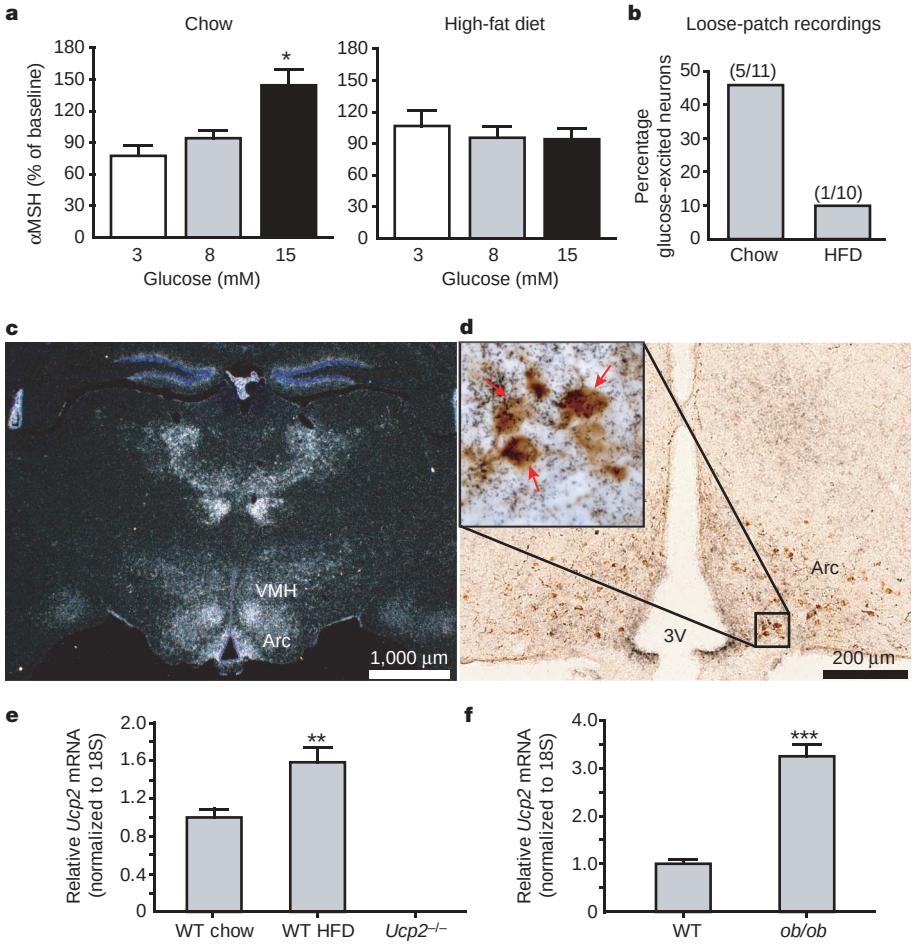


Figure 2 | Glucose-sensing is lost in POMC neurons of mice on a high-fat diet. **a**, Glucose-induced αMSH release from hypothalamic slices of wild-type C57BL/6 mice fed chow or a high-fat diet for 20 weeks (mean ± s.e.m.). **b**, Bar chart showing the percentage of POMC neurons activated by 5 mM glucose in loose-patch recordings from POMC-GFP mice fed either chow or a high-fat diet (HFD) for eight weeks. **c**, *In situ* hybridization of *Ucp2* mRNA in wild-type mice (darkfield photomicrograph of ^{35}S-silvergrains). VMH, ventromedial hypothalamus; Arc, arcuate nucleus. **d**, Double immunohistochemistry and *in situ* hybridization detecting β-endorphin protein and *Ucp2* mRNA, respectively, in coronal sections from wild-type mice. Arrows indicate the presence of β-endorphin neurons co-localized with *Ucp2* mRNA. 3V, third ventricle. **e**, **f**, Relative hypothalamic *Ucp2* mRNA expression in wild-type mice on a high-fat diet (**e**, $n = 12$; ± s.e.m.) and *ob/ob* (**f**, $n = 6$; ± s.e.m.) mice. Two asterisks, $P < 0.01$; three asterisks, $P < 0.001$ compared to wild type (WT).

excited by genipin are the same population of neurons that are also excited by glucose, we performed a series of experiments in the loose-patch mode in which neurons were sequentially tested for responses to glucose and genipin. There was a high degree of concordance between glucose sensing and activation by genipin—if a neuron was activated by glucose, it was usually excited by genipin (Fig. 3a), and if a neuron was glucose-insensitive, it was rarely excited by genipin (Fig. 3b). For 16 out of 22 glucose-excited POMC neurons, genipin significantly increased firing rates (fold increase 2.31 ± 0.30). For 12 out of 13 glucose-insensitive POMC neurons, genipin was without effect (fold increase 1.09 ± 0.20).

Further evidence that genipin is relatively specific for glucose-excited neurons is the finding that neuropeptide Y (NPY) neurons in the arcuate nucleus, which are inhibited by glucose[24], are not excited by genipin (see Supplementary Information). Of note, the ability of genipin to excite POMC neurons was impaired in *Ucp2*$^{-/-}$ mice (Fig. 3c) and in POMC-mut-Kir6.2 mice (Fig. 3d, e). For 11 out of 12 glucose-excited POMC neurons from *Ucp2*$^{-/-}$ mice and 13 out of 14 POMC neurons from POMC-mut-Kir6.2 mice, genipin was without effect (fold increase for *Ucp2*$^{-/-}$ mice, 1.10 ± 0.13; for POMC-mut-Kir6.2 mice, 1.15 ± 0.18). Genipin did increase the firing rate of 1 out of 12 glucose-excited *Ucp2*$^{-/-}$ POMC neurons (1.59-fold increase) and of 1 out of 14 mut-Kir6.2 POMC neurons (1.54-fold increase). However, these increases were much less than typically observed for glucose-excited wild-type POMC neurons (2.30-fold increase). These findings demonstrate that genipin excites POMC neurons by inhibition of UCP2 and subsequent ATP-induced closure of K_{ATP} channels.

To test the hypothesis that increased UCP2 activity mediates loss of glucose sensing in POMC neurons of mice fed a high-fat diet, we assessed whether acute pharmacological inhibition or genetic deletion of UCP2 would reverse or prevent (respectively) loss of glucose sensing. Wild-type and *Ucp2*$^{-/-}$ mice were placed on a high-fat diet for 20 weeks, and hypothalamic slices were tested for release of αMSH in response to glucose. Both wild-type and *Ucp2*$^{-/-}$ mice had increased body weight compared with chow-fed controls. There were no differences in body weight between wild-type or *Ucp2*$^{-/-}$ mice fed on chow (25.1 ± 0.75 g compared to 26.5 ± 0.69 g) or on a high-fat diet (33.2 ± 1.43 g compared to 33.6 ± 1.51 g). As shown in Fig. 4a and observed previously (Fig. 2a), glucose-stimulated release of αMSH was lost in hypothalamic slices from wild-type mice on a high-fat diet. This defective response was fully restored by the acute addition of genipin (Fig. 4a, right panel). *Ucp2*$^{-/-}$ mice were completely protected from diet-induced loss of glucose sensing in POMC neurons (Fig. 4b, right panel). These findings, which are analogous to previous observations in pancreatic β-cells[8,19], indicate that increased UCP2 activity is causally linked to loss of glucose sensing in POMC neurons induced by a high-fat diet.

We suggest several conclusions from the present study. First, we have shown that glucose sensing in POMC neurons has an important role in controlling systemic glucose homeostasis. Second, glucose sensing in these neurons is lost with obesity linked to a high-fat diet. Finally, UCP2 is involved in this loss of glucose sensing, perhaps by decreasing ATP production in POMC neurons. As POMC neurons represent only a fraction of all glucose-excited neurons in the brain (which include melanin-concentrating hormone (MCH) neurons in the lateral hypothalamus[25], neurons in the ventromedial

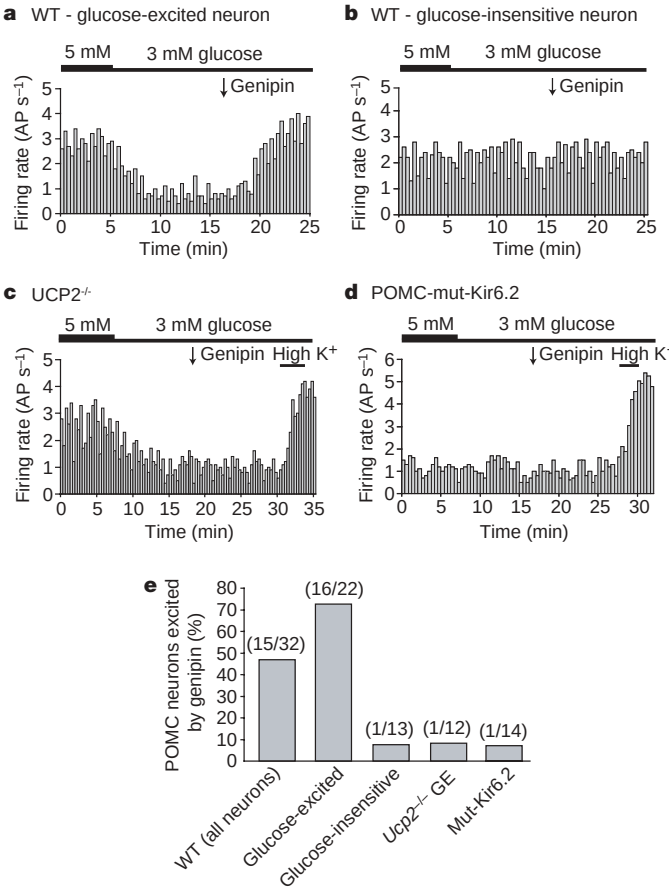


Figure 3 | Genipin activates glucose-excited POMC neurons. Representative time course of firing rates of loose-patch recordings on POMC neurons from wild-type (WT) (**a**, **b**), *Ucp2*$^{-/-}$ (**c**), and POMC-mut-Kir6.2 (**d**) mice. Recordings were made for 5–10 min in aCSF solution containing 5 mM glucose. Once stable firing rates were observed, the recording chamber was perfused with aCSF solution containing 3 mM glucose for 5–10 min, and then genipin (20 μM) was added as indicated by the arrow. **a**, A glucose-excited POMC neuron activated by genipin, representative of 16 out of 22 glucose-excited neurons recorded. **b**, A glucose-insensitive POMC neuron not activated by genipin, representative of 12 out of 13 glucose-insensitive neurons recorded. **c**, A glucose-excited POMC *Ucp2*$^{-/-}$ neuron not activated by genipin, representative of 11 out of 12 glucose-excited *Ucp2*$^{-/-}$ neurons recorded. **d**, A POMC-mut-Kir6.2 neuron not activated by genipin, representing 13 out of 14 neurons recorded. **e**, Bar chart showing the percentage of neurons activated by genipin.

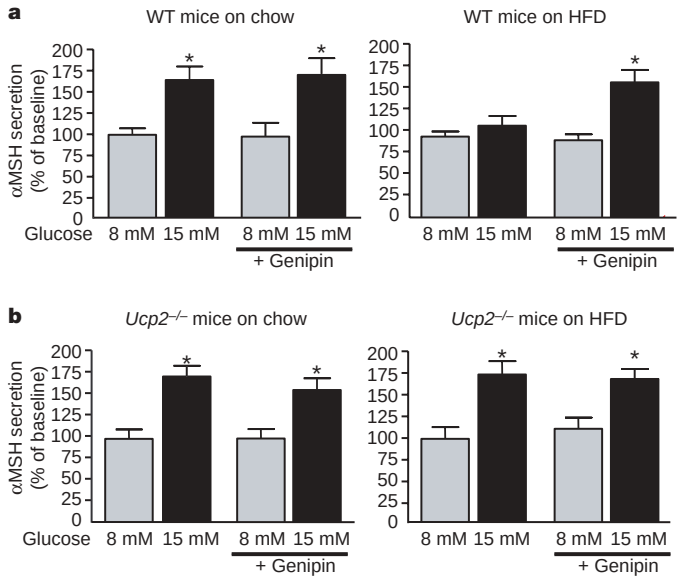


Figure 4 | Acute inhibition or genetic deletion of UCP2 restores or prevents loss of glucose sensing in POMC neurons as a result of obesity induced by a high-fat diet. **a**, **b**, αMSH secretion from hypothalamic slices from wild-type (**a**, WT) and *Ucp2*$^{-/-}$ (**b**) mice in response to glucose, with or without genipin (20 μM). Data are presented as mean ± s.e.m., $n = 6$ mice for each experimental condition. Asterisk, $P < 0.05$.

hypothalamus[2,26], and neurons in the hindbrain[27]), we suggest that UCP2-mediated loss of glucose sensing in glucose-excited neurons could be an important pathogenic component of type 2 diabetes.

METHODS SUMMARY

Transgenic POMC-GFP and POMC-mut-Kir6.2 mice were generated by insertion of humanized renilla (hr)GFP or a Kir6.2 mutant (mut-Kir6.2, Kir[Δ2–30,K185Q]–GFP) (ref. 7) into a POMC bacterial artificial chromosome (BAC) genomic clone, respectively, as described previously[9]. For electrophysiological studies, brain slices were prepared from young adult mice (4–7 weeks old or 12 weeks old for high-fat-diet studies) as described previously[28]. Bath perfusion was used for the addition of glucose (3 mM or 5 mM), genipin (20 μM) or leptin (100 nM) during recording. For αMSH release assays, 2-mm thick hypothalamic slices were incubated in artificial cerebrospinal fluid (aCSF) containing 8 mM glucose (baseline) for 45 min, followed by a 45-min incubation in aCSF containing glucose (as indicated) or genipin (20 μM) (treatment). Tissue viability was verified by a further incubation in aCSF containing KCl, and secreted αMSH was measured by radioimmunoassay as described previously[11]. For glucose tolerance testing, eight-week-old male littermates were fasted overnight before administration of an intraperitoneal (i.p.) glucose load (1 g kg^{-1} body weight). For immunohistochemistry, coronal mouse brain sections (25-μm thick) were prepared as described previously[29] and immunostained with anti-GFP IgG and/or anti-β-endorphin IgG. *In situ* hybridization for *Ucp2* was performed using a ^{35}S-labelled riboprobe complementary to exons 3 and 4 of mouse *Ucp2*. Quantitative PCR was performed on 0.5 ng of hypothalamic cDNA using primers and dual-labelled probes (5′-FAM and 3′-TAMRA) complementary to mouse *Ucp2*, and normalized to levels of the housekeeping gene 18S ribosomal RNA.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 17 July 2007.
Published online 29 August 2007.

1. Ashcroft, F. M., Harrison, D. E. & Ashcroft, S. J. Glucose induces closure of single potassium channels in isolated rat pancreatic beta-cells. *Nature* **312,** 446–448 (1984).
2. Ashford, M. L., Boden, P. R. & Treherne, J. M. Glucose-induced excitation of hypothalamic neurones is mediated by ATP-sensitive K+ channels. *Pflugers Arch.* **415,** 479–483 (1990).
3. Miki, T. *et al.* ATP-sensitive K+ channels in the hypothalamus are essential for the maintenance of glucose homeostasis. *Nature Neurosci.* **4,** 507–512 (2001).
4. Kang, L., Routh, V. H., Kuzhikandathil, E. V., Gaspers, L. D. & Levin, B. E. Physiological and molecular characteristics of rat hypothalamic ventromedial nucleus glucosensing neurons. *Diabetes* **53,** 549–559 (2004).
5. Ibrahim, N. *et al.* Hypothalamic proopiomelanocortin neurons are glucose responsive and express K_{ATP} channels. *Endocrinology* **144,** 1331–1340 (2003).
6. Koster, J. C., Marshall, B. A., Ensor, N., Corbett, J. A. & Nichols, C. G. Targeted overactivity of β cell K_{ATP} channels induces profound neonatal diabetes. *Cell* **100,** 645–654 (2000).
7. Koster, J. C., Sha, Q., Shyng, S. & Nichols, C. G. ATP inhibition of K_{ATP} channels: control of nucleotide sensitivity by the N-terminal domain of the Kir6.2 subunit. *J. Physiol. (Lond.)* **515,** 19–30 (1999).
8. Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, β cell dysfunction, and type 2 diabetes. *Cell* **105,** 745–755 (2001).
9. Balthasar, N. *et al.* Leptin receptor signaling in POMC neurons is required for normal body weight homeostasis. *Neuron* **42,** 983–991 (2004).
10. Silver, I. A. & Erecinska, M. Extracellular glucose concentration in mammalian brain: continuous monitoring of changes during increased neuronal activity and upon limitation in oxygen supply in normo-, hypo-, and hyperglycemic animals. *J. Neurosci.* **14,** 5068–5076 (1994).
11. Enriori, P. J. *et al.* Diet-induced obesity causes severe but reversible leptin resistance in arcuate melanocortin neurons. *Cell Metab.* **5,** 181–194 (2007).
12. Poitout, V. & Robertson, R. P. Minireview: Secondary β-cell failure in type 2 diabetes–a convergence of glucotoxicity and lipotoxicity. *Endocrinology* **143,** 339–342 (2002).
13. Lowell, B. B. & Shulman, G. I. Mitochondrial dysfunction and type 2 diabetes. *Science* **307,** 384–387 (2005).
14. Echtay, K. S. *et al.* Superoxide activates mitochondrial uncoupling proteins. *Nature* **415,** 96–99 (2002).
15. Krauss, S., Zhang, C. Y. & Lowell, B. B. A significant portion of mitochondrial proton leak in intact thymocytes depends on expression of UCP2. *Proc. Natl Acad. Sci. USA* **99,** 118–122 (2002).
16. Laybutt, D. R. *et al.* Genetic regulation of metabolic pathways in β-cells disrupted by hyperglycemia. *J. Biol. Chem.* **277,** 10912–10921 (2002).
17. Krauss, S. *et al.* Superoxide-mediated activation of uncoupling protein 2 causes pancreatic β cell dysfunction. *J. Clin. Invest.* **112,** 1831–1842 (2003).
18. Joseph, J. W. *et al.* Uncoupling protein 2 knockout mice have enhanced insulin secretory capacity after a high-fat diet. *Diabetes* **51,** 3211–3219 (2002).
19. Zhang, C. Y. *et al.* Genipin inhibits UCP2-mediated proton leak and acutely reverses obesity- and high glucose-induced β cell dysfunction in isolated pancreatic islets. *Cell Metab.* **3,** 417–427 (2006).
20. Joseph, J. W. *et al.* Free fatty acid-induced β -cell defects are dependent on uncoupling protein 2 expression. *J. Biol. Chem.* **279,** 51049–51056 (2004).
21. Horvath, T. L. *et al.* Brain uncoupling protein 2: uncoupled neuronal mitochondria predict thermal synapses in homeostatic centers. *J. Neurosci.* **19,** 10417–10427 (1999).
22. Richard, D., Clavel, S., Huang, Q., Sanchis, D. & Ricquier, D. Uncoupling protein 2 in the brain: distribution and function. *Biochem. Soc. Trans.* **29,** 812–817 (2001).
23. Richard, D. *et al.* Distribution of the uncoupling protein 2 mRNA in the mouse brain. *J. Comp. Neurol.* **397,** 549–560 (1998).
24. Mountjoy, P. D., Bailey, S. J. & Rutter, G. A. Inhibition by glucose or leptin of hypothalamic neurons expressing neuropeptide Y requires changes in AMP-activated protein kinase activity. *Diabetologia* **50,** 168–177 (2007).
25. Burdakov, D., Gerasimenko, O. & Verkhratsky, A. Physiological changes in glucose differentially modulate the excitability of hypothalamic melanin-concentrating hormone and orexin neurons *in situ*. *J. Neurosci.* **25,** 2429–2433 (2005).
26. Routh, V. H. Glucosensing neurons in the ventromedial hypothalamic nucleus (VMN) and hypoglycemia-associated autonomic failure (HAAF). *Diabetes Metab. Res. Rev.* **19,** 348–356 (2003).
27. Dallaporta, M., Perrin, J. & Orsini, J. C. Involvement of adenosine triphosphate-sensitive K^+ channels in glucose-sensing in the rat solitary tract nucleus. *Neurosci. Lett.* **278,** 77–80 (2000).
28. Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411,** 480–484 (2001).
29. Dhillon, H. *et al.* Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron* **49,** 191–203 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank J. Koster and C. Nichols for donation of the Kir6.2[ΔN2–30,K185Q]–GFP construct; Z. Yang, L. Christiansen, M. Kramer and S. Skowronek for animal care and technical assistance; and B. Bean for discussions, guidance with electrophysiological experiments and for critical reading of this manuscript. This work was supported by NIH grants (B.B.L., J.K.E., M.A.C., P.J.E.); a Smith Family Pinnacle Award from the American Diabetes Association (J.K.E.); a National Natural Science Foundation of China Outstanding Young Scientist Award; the National Basic Research Program of China (973 Program); the '111' Project; and the Natural Science Foundation of Jiangsu Province (C.-Y.Z.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.A.C. (cowleym@ohsu.edu) or B.B.L. (blowell@bidmc.harvard.edu).

METHODS

Mice. For generation of POMC-mut-Kir6.2 mice, the mut-Kir6.2 cassette (Kir[Δ2–30,K185Q]–GFP)[7] was inserted into a POMC BAC genomic clone so that the ATG codon replaced that of POMC, as described previously[9]. POMC-mut-Kir6.2 BAC DNA was prepared using a commercially available kit (Qiagen) and microinjected into pronuclei of fertilized one-cell-stage embryos of FVB mice (Jackson Laboratories), resulting in the generation of two POMC-mut-Kir6.2 lines that were maintained on an FVB inbred background. POMC-GFP and NPY-GFP mice were generated by insertion of hrGFP into a POMC or NPY BAC, respectively, as described above. *Ucp2*$^{-/-}$ mice were used as described previously[8]. To generate POMC-GFP;*Ucp2*$^{-/-}$ mice, heterozygous POMC-GFP transgenic mice were crossed with heterozygous *Ucp2*$^{-/-}$ mice. For high-fat diet feeding experiments, mice were placed on a high-fat rodent diet (45% kcal from fat; Research Diets Inc; D12451) at four weeks of age for a total of 20 weeks (or 8 weeks for electrophysiological studies).

Electrophysiology. Brain slices were prepared from young adult mice (4–7 weeks old or 12 weeks old for high-fat diet studies) as described previously[28]. Briefly, mice were anesthetized with Isoflurane before decapitation and removal of the entire brain. Brains were immediately submerged in ice-cold, carbogen-saturated (95% O_2, 5% CO_2) high-sucrose solution (238 mM sucrose, 26 mM $NaHCO_3$, 2.5 mM KCl, 1.0 mM NaH_2PO_4, 5.0 mM $MgCl_2$, 1.0 mM $CaCl_2$, 11 mM D-glucose). Coronal sections (200-μm thick) were cut with a Leica VT1000S Vibratome and incubated in oxygenated recording aCSF (126 mM NaCl, 21.4 mM $NaHCO_3$, 2.5 mM KCl, 1.2 mM NaH_2PO_4, 1.2 mM $MgCl_2$, 2.4 mM $CaCl_2$, 10 mM D-glucose) at room temperature (21–24 °C) for at least 1 h before recording. Slices were transferred to the recording chamber and bathed in oxygenated recording aCSF heated to approximately 34 °C, at a flow rate of approximately 2 ml min^{-1}. GFP-positive neurons were visualized using epifluorescence imaging and patched under infrared differential interference contrast (IR-DIC) optics. Recordings were made using a MultiClamp 700B amplifier using pClamp 9.0 software (both Axon Instruments).

Bath perfusion was used for the addition of glucose, genipin and leptin during recording. Equal molar NaCl was replaced with 56 mM KCl to prepare 'high-K^+ aCSF'. The pipette solution for loose-patch recording contained 150 mM NaCl, 3.5 mM KCl, 10 mM HEPES, 10 mM glucose, 2.5 mM $CaCl_2$, 1.3 mM $MgCl_2$ (pH 7.3). Loose-patch recordings were performed in $I = 0$ current-clamp mode, which maintains an average 0 pA holding current[30]. Seal resistance was in the range of 8–30 MΩ and was checked during the recordings; cells that showed deviations in seal resistance were not included in the data analysis. Recordings that showed no recovery in spike activity after return to 5 mM glucose were also excluded from data analysis. Firing rate averaged for every 20 s was taken as one data point; 9–18 data points taken from the last 3–6 min of each experimental condition (5 mM or 3 mM glucose, with or without addition of drugs) were compared using unpaired t-tests, with $P < 0.05$ considered a statistically significant change.

Static incubation of hypothalamic explants. Mice were killed by decapitation and whole brains removed immediately. A 2-mm slice was prepared using a vibrating microtome (Leica VS 100) taken from the base of the brain to include the paraventricular nucleus (PVH) and arcuate nucleus. Each hypothalamic slice was treated separately and incubated in artificial cerebrospinal fluid (aCSF: 126 mM NaCl, 0.09 mM Na_2HPO, 6 mM KCl, 4 mM $CaCl_2$, 0.09 mM $MgSO_4$, 20 mM $NaHCO_3$, 8 mM glucose, 0.18 mg ml^{-1} ascorbic acid, 0.6 trypsin inhibitor units (TIU) aprotinin ml^{-1}), pre-equilibrated with 95% O_2 and 5% CO_2 at 37 °C for 1 h. Slices were then incubated for 45 min in 700 μl aCSF containing 8 mM glucose (baseline) followed by a 45-min incubation in aCSF containing 0.1, 0.5, 1.0, 5.0 or 10 mM glucose for dose–response experiments, or glucose (3, 8 or 15 mM) or genipin (20 μM) for feeding experiments (high-fat diet versus chow). Finally, the viability of the tissue was verified by a 45-min incubation in aCSF containing 56 mM KCl. At the end of each incubation period, supernatants were removed and tested for αMSH release by radioimmunoassay as previously described[11]. αMSH secretion from each individual hypothalamus was normalized to the amount of αMSH secreted during the baseline period (8 mM glucose). Only hypothalami that showed a 300% secretion over baseline in response to 56 mM KCl were used in the analysis. Individual experiments were repeated a total of four times.

Glucose tolerance testing. Mice were fasted overnight (16 h) and injected intraperitoneally with a 20% (w/v) glucose solution at 1 g kg^{-1} body weight. Blood glucose levels were measured before and 15, 30, 45, 90, 120 and 180 min after glucose injection.

Immunohistochemistry. Coronal mouse brain sections (25 μm) were washed in PBS six times before blocking in 0.25% Triton X-100 in PBS containing 3% (w/v) normal donkey serum (PBT-azide) for 2 h. Sections were incubated overnight in rabbit anti-β-endorphin IgG (Peninsula Laboratories Inc; 1:5,000) in PBT-azide containing 3% (w/v) normal donkey serum at room temperature (21–24 °C), followed by a 2-h incubation in Cy3-conjugated donkey anti-rabbit IgG (ImmunoResearch Laboratories; 1:500). After three washes in PBS, sections were blocked in PBT-azide containing 3% (w/v) normal donkey serum and incubated overnight at room temperature in chicken anti-GFP IgG (Upstate Laboratories; 1:5,000) and Cy2-conjugated streptavidin (Jackson ImmunoResearch Laboratories; 1:500) for 1 h. Sections were mounted onto Superfrost slides and visualized on an inverted microscope with a digital camera (Axioscope, Zeiss). A total of four brains for each POMC-mut-Kir6.2 transgenic line were processed and imaged. Each brain was divided into five serial groups, with each series containing a comparable representation of the entire brain. Four out of the five series were used to count the total number of both GFP- and β-endorphin-positive neurons. For each of the two transgenic lines that were used for further study (lines 1 and 2), the percentage of β-endorphin-positive cells that co-expressed GFP was >95% and the percentage of GFP-positive cells that were negative for β-endorphin expression was <1%.

***Ucp2 in situ* hybridization.** A *Ucp2* riboprobe was generated by PCR amplification of a 527-base pair (bp) DNA fragment complementary to exons 3 and 4 of *Ucp2* from mouse brain cDNA. This amplicon was subcloned into pcR4-TOPO vector (Invitrogen), and a ^{35}S-labelled cRNA probe generated as described previously[29]. For β-endorphin immunohistochemistry, sections processed for *in situ* hybridization were washed twice in PBS and pre-treated with 0.3% (v/v) hydrogen peroxide in PBS for 30 min, then incubated in 3% normal donkey serum in PBT-azide for 2 h. Sections were then incubated overnight with rabbit anti-β-endorphin primary antiserum (Peninsula Laboratories; 1:4,000) in PBT-azide. Sections were washed in PBS six times before a 2-h incubation in biotinylated donkey anti-rabbit IgG (Jackson ImmunoResearch Laboratories; 1:1,000). Sections were then washed three times in PBS and incubated with avidin–biotin complex (Vectastain Elite ABC Kit, Vector Laboratories; 1:500) in PBS for 1 h. Finally, sections were washed three times in PBS and incubated in 0.04% diaminobenzidine tetrahydrochloride (DAB; Sigma) and 0.01% hydrogen peroxide in PBS. The DAB reaction was quenched by two washes with PBS.

Quantitative real-time PCR. Hypothalami were removed and snap-frozen in liquid nitrogen before isolation of total RNA using RNA STAT-60 (Tel-Test Inc) according to the manufacturer's instructions. Samples were then treated with DNA-Free (Ambion) to remove any contaminating genomic DNA, and complementary DNA was synthesized from 1 μg of total RNA using Superscript III FirstStrand cDNA Synthesis Kit (Invitrogen). *Ucp2* was amplified from 0.5 ng of reverse-transcribed total RNA using Taqman Universal PCR Mastermix (Applied Biosystems) with *Ucp2* sense and antisense primers, and a dual-labelled probe (5'-FAM, 3'-TAMRA) (Applied Biosystems; assay on demand Mn00495907_g1). Standard curves were constructed by amplifying serial dilutions of cDNA (5 ng to 0.32 pg) and plotting cycle threshold (CT) values as a function of starting reverse-transcribed RNA. mRNA expression of *Ucp2* was normalized to levels of the 18S ribosomal RNA housekeeping gene.

30. Perkins, K. L. Cell-attached voltage-clamp and current-clamp recording and stimulation techniques in brain slices. *J. Neurosci. Methods* **154**, 1–18 (2006).

LETTERS

The structural basis of yeast prion strain variants

Brandon H. Toyama[1], Mark J. S. Kelly[2], John D. Gross[2] & Jonathan S. Weissman[1]

Among the many surprises to arise from studies of prion biology, perhaps the most unexpected is the strain phenomenon whereby a single protein can misfold into structurally distinct, infectious states that cause distinguishable phenotypes[1–3]. Similarly, proteins can adopt a spectrum of conformations in non-infectious diseases of protein folding; some are toxic and others are well tolerated[4]. However, our understanding of the structural differences underlying prion strains and how these differences alter their physiological impact remains limited. Here we use a combination of solution NMR, amide hydrogen/deuterium (H/D) exchange and mutagenesis to study the structural differences between two strain conformations, termed Sc4 and Sc37 (ref. 5), of the yeast Sup35 prion. We find that these two strains have an overlapping amyloid core spanning most of the Gln/Asn-rich first 40 amino acids that is highly protected from H/D exchange and very sensitive to mutation. These features indicate that the cores are composed of tightly packed β-sheets possibly resembling 'steric zipper' structures revealed by X-ray crystallography of Sup35-derived peptides[6,7]. The stable structure is greatly expanded in the Sc37 conformation to encompass the first 70 amino acids, revealing why this strain shows increased fibre stability and decreased ability to undergo chaperone-mediated replication[8]. Our findings establish that prion strains involve large-scale conformational differences and provide a structural basis for understanding a broad range of functional studies, including how conformational changes alter the physiological impact of prion strains.

Sup35, the protein determinant of the yeast prion [PSI^+], is uniquely well suited to the analysis of the structural basis of prion strains[3]. Pure Sup35 spontaneously forms self-seeding, β-sheet-rich amyloid fibres *in vitro*[9]. Introduction of these fibres into yeast causes stable conversion to the [PSI^+] state, thus establishing their infectious (prion) nature[5,10,11]. Sup35 can adopt a variety of fibre conformations *in vitro*, which lead to clearly distinguishable prion strain variants when introduced into [psi^-] cells. For example, SupNM, a region of Sup35 encoding its prion function and composed of a Gln/Asn-rich N-terminal domain (N, amino-acid residues 1–123) and a highly charged middle domain (M, residues 124–253), can adopt distinct fibre conformations when polymerization is initiated at 4 °C and at 37 °C. Introduction of these conformations, termed Sc4 and Sc37, into yeast induces 'strong' and 'weak' *in vivo* prion strain phenotypes, respectively[5]. Thus, the heritable differences in these strain variants are 'encoded' by the conformational differences between Sc4 and Sc37.

Although a battery of biophysical techniques have been employed to investigate Sup35 prion structures, no consensus view has emerged. One set of approaches introduced cysteine residues and inferred structure either by measuring side-chain reactivity or by incorporating biophysical probes such as paramagnetic spin labels or pyrene fluorophores[5,12]. One such study suggested that there is stable structure roughly comprising residues 31–86 and 21–121 for Sc4 and Sc37, respectively, with flanking sequences not being part of this amyloid core[12]. Close intermolecular contacts between monomers seemed to be limited to short regions involving 'head-to-head' (residues 25–38) and 'tail-to-tail' (residues 91–106) interactions. An alternative view emerged from X-ray crystallography of the amyloid-like structure of the GNNQQNY peptide corresponding to residues 7–13 of Sup35 (ref. 6), a region outside the amyloid core suggested by the above study. Unlike the proposed head-to-head/tail-to-tail assembly, GNNQQNY formed extensive intermolecular contacts through in-register parallel β-sheets that were stabilized by stacking side-chain amide hydrogen bonds between analogous asparagine and glutamine residues. Two sheets then came together face-to-face to form a water-free 'dry' interface in which opposing side chains interdigitated with each other, creating a compact 'steric zipper'. Caveats exist for both approaches: the X-ray crystallography analyses examined peptides outside the context of the prion domain, and the cysteine studies used side-chain modifications that might have perturbed the structure and stability of the fibres. Indeed, the extremely tight packing seen in the dry interfaces would probably be intolerant of side-chain alterations. More recently, solid-state NMR was used to probe for secondary structure in SupNM fibres in which a subset of amino acids were isotopically labelled[13]. Although it was not possible to obtain sequence-specific assignments, seven of eight leucine residues found in SupNM were suggested to form in-register β-sheets. Because all except one leucine residue are in the M domain, it was proposed that much of the M domain was structured in the solid state. This conclusion was surprising because the M domain is not absolutely required for prion function[14] and was thought to be largely disordered in fibres in solution[5,12]. These conflicting perspectives emphasize the need for residue-specific structural information on full-length, unmodified SupNM fibres in solution.

We first examined SupNM fibre structure by solution NMR. Large complexes (that is, more than 100 kDa) are generally not amenable to solution NMR, as a result of line broadening, unless specialized spectroscopic techniques such as TROSY are employed[15]. However, highly mobile regions within large complexes can sometimes be detected by using standard heteronuclear single-quantum coherence (HSQC) pulse sequences. Indeed, substantial portions of SupNM seem to be disordered in fibre forms, because ^{15}N-HSQC spectra from uniformly ^{15}N-labelled SupNM fibres revealed multiple robust peaks (Fig. 1a). To assess which regions of SupNM are flexible, we employed a strategy similar to that used in the solid-state NMR study and specifically ^{15}N-labelled the leucine residues. In dimethylsulphoxide (DMSO), in which SupNM is monomeric, seven of eight potential leucine peaks were observed, demonstrating the specificity of the labelling (Supplementary Fig. 1). NMR spectra of ^{15}N-Leu-SupNM fibres contained at least four robust peaks with a smaller additional peak (Fig. 1b). Because all except one of the leucines are in the M domain (Fig. 1c), these data indicate that, in solution, large regions of the M domain remain highly mobile even in the fibre form, a view substantiated by the experiments below.

[1]Howard Hughes Medical Institute, Department of Cellular and Molecular Pharmacology, [2]Department of Pharmaceutical Chemistry, University of California San Francisco and California Institute for Quantitative Biomedical Research, San Francisco, California 94158-2542, USA.

To obtain residue-specific information on the Sc4 and Sc37 conformations, we measured quenched hydrogen/deuterium (H/D) exchange by using NMR. H/D exchange is a sensitive probe of structure because hydrogen bonds inhibit the exchange of amide protons with water. We employed a recent protocol[16] that has enabled the structural analysis of a variety of amyloid fibres[16–20] by exploiting the ability of DMSO to dissolve fibres while preserving their exchange state. DMSO effectively dissolved SupNM fibres, yielding a well-dispersed ^{15}N-HSQC spectrum (Fig. 2a and Supplementary Fig. 3a–c). However, assignment of the ^{15}N-HSQC, a prerequisite for measuring exchange, remained a challenge. SupNM is large (253 residues) compared with other amyloidogenic polypeptides studied[16,18–20], and the N domain has very low complexity; the first 40 residues (Sup1–40) are especially Gln/Asn-rich, and the adjacent residues 41–97 are composed of 5.5 imperfect oligopeptide repeats of the sequence PQGGYQQYN (Supplementary Fig. 2). Through the use of more than 20 three-dimensional heteronuclear NMR experiments on 11 different uniformly and specifically labelled samples, we succeeded in assigning 163 of 215 visible peaks (Fig. 2a and Supplementary Fig. 2) including assignments for 33 of the first 40 residues, several assignments in each oligopeptide repeat, and most of the M domain.

We monitored the extent of H/D exchange on uniformly ^{15}N-labelled Sc4 and Sc37 fibres at neutral pH at five time points ranging from 1 min to 1 week. Additionally, we performed H/D exchange for 1 day on Sc4 and Sc37 fibres specifically labelled with ^{15}N-Glu/Gln (Glx) to obtain data on key residues that had significant overlap in the uniformly labelled samples. Examination of the exchange spectra revealed multiple peaks that strongly resisted exchange in both conformation-independent and conformation-dependent manners (Fig. 2b–d). Three classes of exchange curves predominated (Supplementary Fig. 3d). In the first class, amides completed exchange within 1 min, which corresponded to protection factors (the intrinsic rate of exchange divided by the observed rate) of less than 100. The second class showed only modest exchange even after 1 week, corresponding to highly stable regions with protection factors of at least 10^6. The third class of residues did not follow a single exponential: partial exchange was seen in the early time points, but the remaining fraction remained largely unexchanged by 1 week. Similar behaviour has been observed for other amyloid species[17,19] and may reflect multiple structural environments for these residues within the same fibre (see ref. 17).

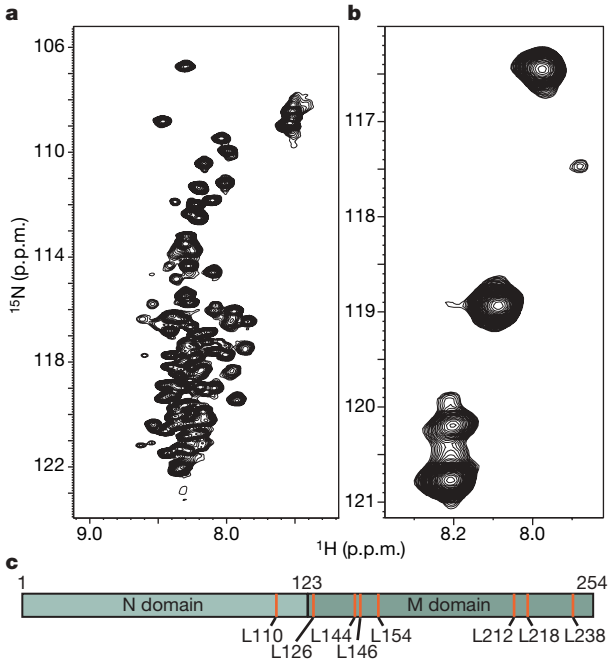


Figure 1 | Solution NMR of SupNM fibres. a, ^{15}N-HSQC spectrum of uniformly ^{15}N-labelled SupNM fibres polymerized into the Sc4 conformation. **b**, ^{15}N-HSQC spectrum of SupNM Sc4 fibres specifically labelled with ^{15}N-leucine. **c**, Diagram showing the distribution of leucine residues in the SupNM sequence.

Our data reveal overlapping core amyloid regions for Sc4 and Sc37 that encompass almost the entire Gln/Asn-rich Sup1–40 subdomain as well as a marked, strain-specific expansion of protected structure into the oligopeptide repeats in the Sc37 fibres (Fig. 2c, d). Both Sc4 and Sc37 conformations showed strong protection for residues 4–37, including the residues containing the GNNQQNY peptide, but had local differential protection patterns. In particular, adjacent residues 38–70, which encompass the first three oligopeptide repeats, showed high levels of protection in Sc37 but completed exchange within the first 1 min in Sc4. Both conformations seemed largely unprotected for residues 76–100, but significant protection was seen for some residues in the 110–128 region. Most of the remaining M domain, including the leucine residues, showed modest protection (residues 150–231 in Supplementary Fig. 4), which is consistent with the solution NMR results.

To test these findings independently, we probed for fibre secondary structure by introducing proline residues throughout SupNM to interfere with β-sheet formation. Analysis of the conformation-specific effects of mutations on fibre formation poses a challenge because Sup35 can adopt multiple amyloid conformations. Hence, mutations detrimental to one fibre form can drive the formation of an alternative conformation rather than preventing fibre polymerization altogether[21]. To circumvent this problem, we monitored the initial rate at which mutant monomers add to the end of preformed wild-type Sc4 and Sc37 fibres by using assays based on thioflavinT (thioT) and atomic force microscopy (AFM)[21,22]. These measurements directly assess the ability of each conformation to accommodate the mutation (Fig. 3a), thereby yielding conformation-specific structural information (Fig. 3b). We analysed a series of 36 individual SupNM proline mutations with locations chosen to overlap with and extend the H/D exchange data (Supplementary Fig. 2). Of these mutations, 16 lie in Sup1–40, another 16 lie in the subsequent 41–100 oligopeptide region, and the last four target residues 101–150. Because the oligopeptide repeats were a major source of missing assignments, two corresponding Gln residues near the centre of each repeat were individually mutated to proline.

The proline mutant analysis showed remarkable agreement with the H/D exchange data. With the exception of the termini, residues within the Sup1–40 subdomain were highly sensitive to mutations for both Sc4 and Sc37 seeds (Fig. 3c). AFM directly confirmed that proline mutations in this region profoundly inhibited the ability of mutant monomers to add to the ends of wild-type fibres (Fig. 3c inset, and Supplementary Fig. 5a). In contrast, mutations within the first three oligopeptide repeats affected growth from Sc37 but not Sc4 seeds, whereas mutations from the fourth repeat up to the M domain had modest effects on growth for either conformation (Fig. 3c).

We extended this mutational analysis to probe tertiary structural contacts, focusing on the NQQQ sequence (residues 21–24) in the centre of the amyloid core. To minimize secondary structure perturbations, residues were individually changed to leucine, which has a similar β-sheet propensity to that of both asparagine and glutamine. Leucine also has a higher intrinsic aggregation propensity[4] and is isosteric to asparagine. Despite the modest nature of the change, we saw strong inhibitory effects of leucine mutations for both Sc4 and Sc37 fibres (Fig. 3d and Supplementary Fig. 5b). This indicates a critical role for the Asn/Gln side chains in stabilizing the amyloid core, possibly through interstrand hydrogen bonds and face-to-face interactions between β-sheets such as those seen in the GNNQQNY peptide structure. Interestingly, Sc4 but not Sc37 displayed an alternating pattern of sensitivity to leucines, possibly indicative of a β-sheet structure with one face buried and the other accessible in

Sc4, but with both faces buried in Sc37 (Fig. 3d). Our data, in conjunction with those from earlier analyses[6,7,13,23], suggest that the Sup1–40 core may resemble the Sup35 peptide structures, with the exception that the dry sheet–sheet interfaces are probably formed by intramolecular interactions between β-sheet faces from different rather than identical Asn/Gln-rich regions. Because some differences in protection patterns within Sup1–40 are observed between Sc4 and Sc37, these two strains may have distinct sheet–sheet interfaces in which the Sc37 arrangement accommodates additional β-sheets from the oligopeptide repeats. Such diverse sheet–sheet interfaces have recently been described in peptide structures[7].

Our finding that the Gln/Asn-rich Sup1–40 subdomain forms an overlapping amyloid core that expands well into the adjacent oligopeptide repeats in a strain-dependent manner provides a structural rationale for several diverse functional studies. First, mutational studies and a recent peptide-mapping analysis indicate that prion formation and specificity of growth are particularly sensitive to changes within Sup1–40 (refs 24–26). Second, the SupNM prion domain has a modular architecture in which Sup1–40 is specifically required for amyloid formation and templated growth, and can be functionally replaced by amyloidogenic stretches of pure polyGln repeats[27]. In addition to a role for a

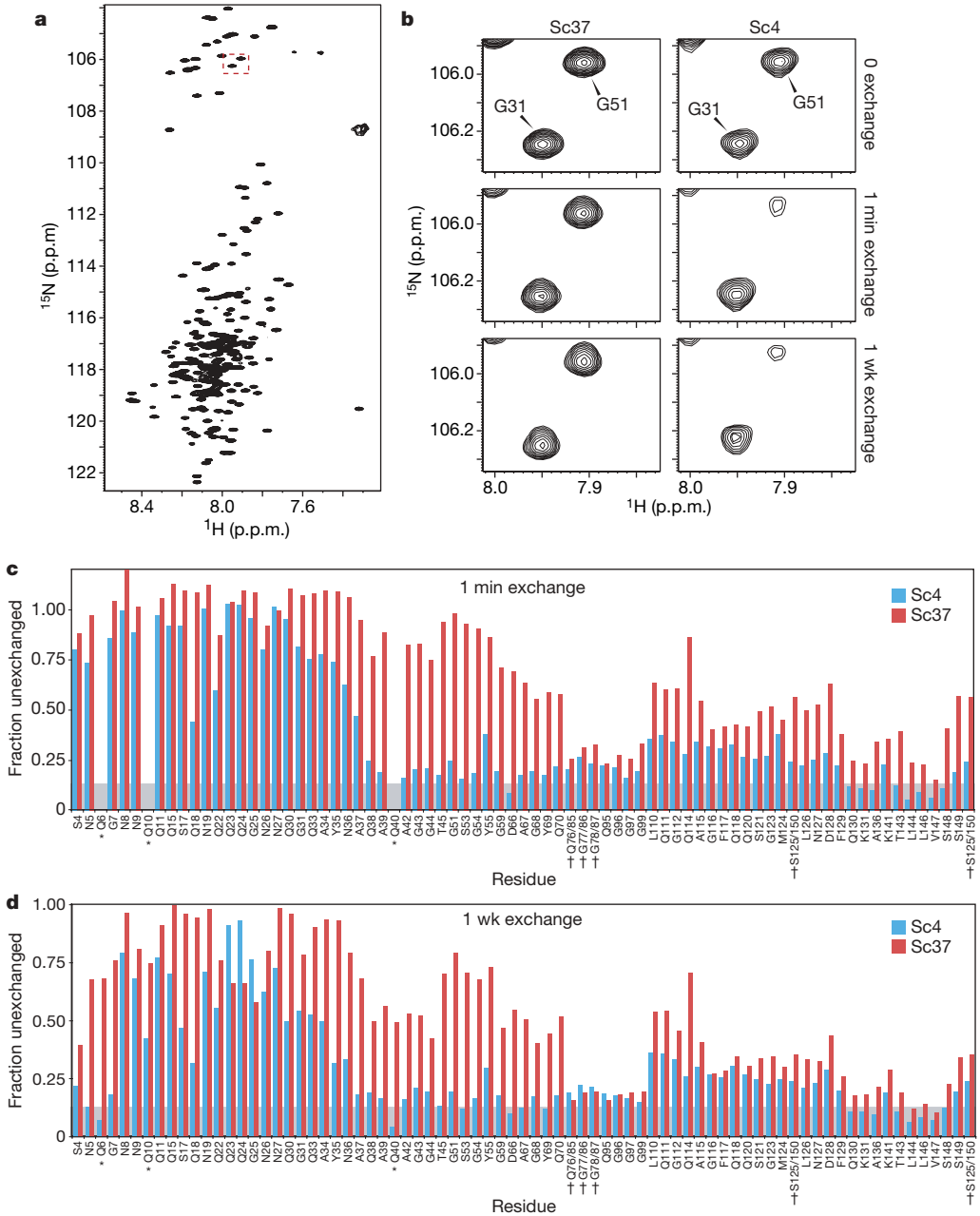


Figure 2 | H/D exchange of Sc4 and Sc37 fibres. a, ^{15}N-HSQC spectrum of uniformly ^{15}N-labelled SupNM fibres dissolved in DMSO dissolution buffer. The red dashed box indicates residues shown in **b**. **b**, H/D exchange performed on the Sc4 and Sc37 conformations. ^{15}N-HSQC spectra for residues Gly 31 and Gly 51 (red dashed box in **a**) for the Sc37 (left panels) and Sc4 (right panels) conformations after 0 min (top panels), 1 min (middle panels) and 1 week (bottom panels) of exchange. **c**, **d**, Intensities for all assigned and unambiguous peaks as a fraction of the non-exchanged intensity after 1 min (**c**) and 1 week (**d**) of exchange on the Sc4 (blue) and Sc37 (red) conformations. Unassigned and ambiguous residues are not displayed. S4 and N5 in the Sc4 conformation have calculated protection factors of 6,478 ($R^2 = 0.88$) and 26,304 ($R^2 = 0.93$) respectively. Residues marked with an asterisk are resolved only in Glx specifically labelled H/D exchange experiments performed with a single 1-day exchange time point. Daggers indicate pairs of residues with overlapping peaks; values therefore represent the combined intensities. The grey bar represents the estimated minimum peak intensity.

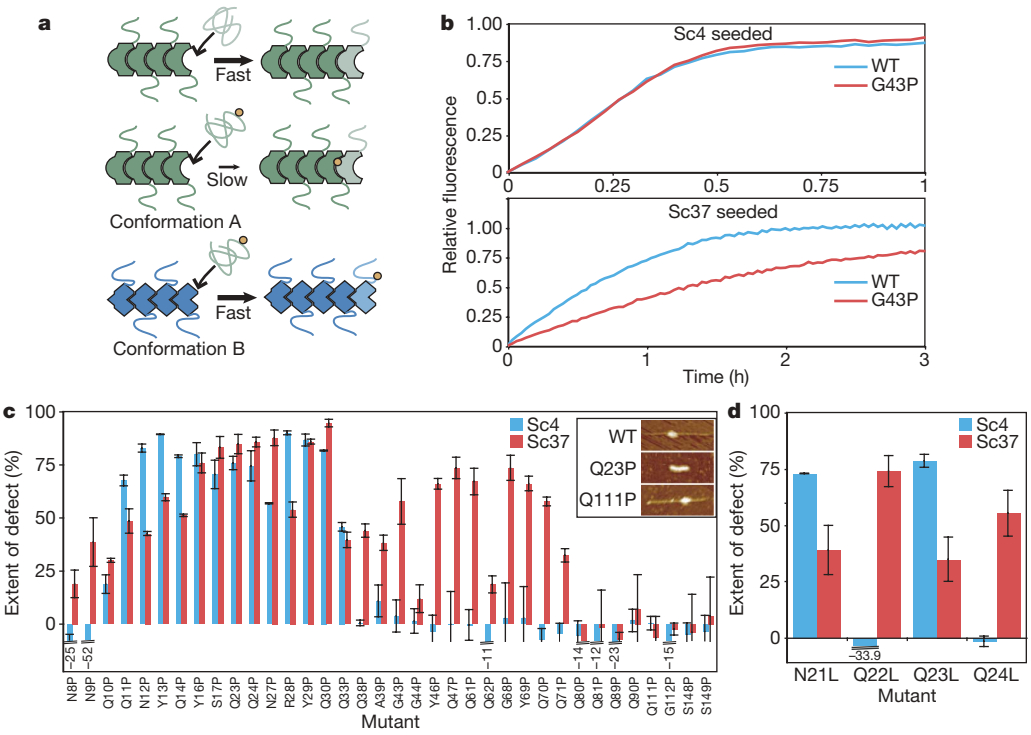


Figure 3 | Mutational analysis of SupNM fibres. a, Diagram of the rationale for analysing the initial rate of polymerization. Here, the introduction of mutations (indicated by the circle) in regions important for fibre structure in conformation A but not in conformation B result in a slow initial rate of polymerization in a conformation-specific manner. **b**, Effect of proline mutants on polymerization as monitored by increase in thioT fluorescence. Wild-type Sc4 (upper panel) and Sc37 (lower panel) seeds were added to monomeric wild-type SupNM (blue curves) and SupNM G43P (red curves) and fibre polymerization was monitored by thioT fluorescence[21]. Data were normalized to initial and final intensities. **c**, Effect of proline mutations on initial rates of polymerization of the Sc4 (blue) and Sc37 (red) conformations. ThioT polymerization assay was performed as described above for all mutants indicated. Data were normalized to initial and final intensities, initial time points were fitted to a line, and the slope (initial rate) was calculated. Values are the percentage by which each indicated mutant's initial rate deviated from the rate of wild-type polymerization, calculated as $100 \times (1 - K_{mut}/K_{wt})$. Error bars indicate s.d. for two or three independent replicates. The inset shows representative pictures of indicated proline mutant growth (thin region) from wild-type (WT) Sc4 seed (thick region) revealed by AFM. **d**, The effect of leucine mutations on initial rates was measured as described above for the indicated leucine mutants of the Sc4 (blue) and Sc37 (red) conformations. Error bars indicate s.d. for two or three independent replicates.

subset of the oligopeptide repeats in stabilizing amyloid structure, these repeats seem to have a distinct and essential function in allowing chaperone-mediated replication of such amyloids, thereby generating new prion seeds. This role in prion division can be replaced by a number of diverse sequences[28,29]. The less stable structure over much of the oligopeptide repeats in addition to their more diverse sequence composition may facilitate recognition of this region by chaperones[3]. Last, the structural differences observed here shed light on how conformations alter *in vivo* phenotypes as well as leading to strain-specific sequence requirements for prion propagation. Ordering of the first three oligopeptide repeats in Sc37 is associated with increased stability of these fibres[5,8] and may also occlude chaperone recognition sites. Taken together, these findings account for the observation that the weaker strain phenotype of Sc37 prions results from a difficulty in generating new prion seeds despite robust fibre growth. Previous studies also found that an increased number of oligopeptide repeats (four versus two) was minimally required to support the propagation of a 'weak' strain as compared with a 'strong' strain[30]. Thus, for both strains, [PSI^+] propagation seems to require at least one oligopeptide repeat outside the highly protected core. The mammalian prion protein, PrP, contains peptide repeat regions reminiscent of Sup35, whose number modulates prion propensity even though they lie outside the protease-resistant prion core[1]. This raises the possibility that these otherwise unrelated prion proteins share a common functional architecture in which an amyloid core mediates templated growth, and less stably structured regions facilitate the generation of new prion seeds.

Note added in proof: Mutation analysis indicated that of the 23 glycines, only the 2 present in the M domain were visible by solution NMR of uniformly labelled fibres, which is consistent with the conclusion that the M domain is highly mobile.

METHODS SUMMARY

SupNM isotope labelling. Labelled SupNM was expressed in *Escherichia coli*.

NMR on fibres. SupNM fibres uniformly labelled with ^{15}N, or specifically labelled with ^{15}N-Leu, were seeded and polymerized at 4 °C as described[5]. Fibres were spun down and resuspended in 1/40 volume of equivalent buffer in 10% (v/v) D_2O (pH 5.5 for uniformly labelled). Standard ^{15}N-HSQC spectra were recorded at 800 MHz. The final protein concentration is estimated as 200 μM. Only a few peaks with chemical shifts typical of glycine were seen in the uniformly labelled sample. The failure to detect all 23 glycine residues in SupNM indicates that the peaks observed were from fibres rather than from a monomer subpopulation.

H/D exchange. H/D exchange was performed on ^{15}N-SupNM seeded fibres largely as described[16], with time points at 0 min, 1 min, 1 h, 1 day and 1 week of exchange. For NMR measurements, fibre pellets were dissolved in dissolution buffer (95% DMSO-d_6, 4.5% D_2O, 0.5% dichloroacetic acid-d_2 (v/v) and 200 μM 2,2-dimethyl-2-silapentane-5-sulphonate sodium salt (DSS) at pH 5.0) and ^{15}N-HSQC correlation spectra were recorded at 800 MHz and 298 K. Dichloroacetic acid-d_2 was purchased from CDN Isotopes. Spectra were recorded in accordance with the strategy described[17]. Additionally, a one-dimensional ^{1}H spectrum was acquired after ^{15}N-HSQC.

Mutagenesis studies. All mutant plasmids were generated, overexpressed and purified as described[5,26]. ThioT assays were performed as described[21]. Sc4 and Sc37 seeded reactions were performed at 4 °C and 25 °C, respectively. Initial rates were calculated from the first 16 min of polymerization. AFM assays were performed as described[22].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 May; accepted 20 July 2007.
Published online 2 September 2007.

1. Cohen, F. E. & Prusiner, S. B. Pathologic conformations of prion proteins. *Annu. Rev. Biochem.* **67,** 793–819 (1998).
2. Derkatch, I. L., Chernoff, Y. O., Kushnirov, V. V., Inge-Vechtomov, S. G. & Liebman, S. W. Genesis and variability of [*PSI*] prion factors in *Saccharomyces cerevisiae*. *Genetics* **144,** 1375–1386 (1996).
3. Tuite, M. F. & Cox, B. S. The [*PSI*+] prion of yeast: a problem of inheritance. *Methods* **39,** 9–22 (2006).
4. Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75,** 333–366 (2006).
5. Tanaka, M., Chien, P., Naber, N., Cooke, R. & Weissman, J. S. Conformational variations in an infectious protein determine prion strain differences. *Nature* **428,** 323–328 (2004).
6. Nelson, R. *et al.* Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435,** 773–778 (2005).
7. Sawaya, M. R. *et al.* Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **447,** 453–457 (2007).
8. Tanaka, M., Collins, S. R., Toyama, B. H. & Weissman, J. S. The physical basis of how prion conformations determine strain phenotypes. *Nature* **442,** 585–589 (2006).
9. Glover, J. R. *et al.* Self-seeded fibers formed by Sup35, the protein determinant of [*PSI*+], a heritable prion-like factor of *S. cerevisiae*. *Cell* **89,** 811–819 (1997).
10. Sparrer, H. E., Santoso, A., Szoka, F. C. Jr & Weissman, J. S. Evidence for the prion hypothesis: induction of the yeast [*PSI*+] factor by *in vitro*-converted Sup35 protein. *Science* **289,** 595–599 (2000).
11. King, C. Y. & Diaz-Avalos, R. Protein-only transmission of three yeast prion strains. *Nature* **428,** 319–323 (2004).
12. Krishnan, R. & Lindquist, S. L. Structural insights into a yeast prion illuminate nucleation and strain diversity. *Nature* **435,** 765–772 (2005).
13. Shewmaker, F., Wickner, R. B. & Tycko, R. Amyloid of the prion domain of Sup35p has an in-register parallel β-sheet structure. *Proc. Natl Acad. Sci. USA* **103,** 19754–19759 (2006).
14. Liu, J. J., Sondheimer, N. & Lindquist, S. L. Changes in the middle region of Sup35 profoundly alter the nature of epigenetic inheritance for the yeast prion. *Proc. Natl Acad. Sci. USA* [*PSI*+] **99** (suppl. 4), 16446–16453 (2002).
15. Flaux, J., Bertelsen, E. B., Horwich, A. L. & Wuthrich, K. NMR analysis of a 900K GroEL–GroES complex. *Nature* **418,** 207–211 (2002).
16. Hoshino, M. *et al.* Mapping the core of the β_2-microglobulin amyloid fibril by H/D exchange. *Nature Struct. Biol.* **9,** 332–336 (2002).
17. Yamaguchi, K. *et al.* Core and heterogeneity of β_2-microglobulin amyloid fibrils as revealed by H/D exchange. *J. Mol. Biol.* **338,** 559–571 (2004).
18. Ritter, C. *et al.* Correlation of structural elements and infectivity of the HET-s prion. *Nature* **435,** 844–848 (2005).
19. Luhrs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1–42) fibrils. *Proc. Natl Acad. Sci. USA* **102,** 17342–17347 (2005).
20. Carulla, N. *et al.* Molecular recycling within amyloid fibrils. *Nature* **436,** 554–558 (2005).
21. Chien, P., DePace, A. H., Collins, S. R. & Weissman, J. S. Generation of prion transmission barriers by mutational control of amyloid conformations. *Nature* **424,** 948–951 (2003).
22. DePace, A. H. & Weissman, J. S. Origins and kinetic consequences of diversity in Sup35 yeast prion fibers. *Nature Struct. Biol.* **9,** 389–396 (2002).
23. Ross, E. D., Edskes, H. K., Terry, M. J. & Wickner, R. B. Primary sequence independence for prion formation. *Proc. Natl Acad. Sci. USA* **102,** 12825–12830 (2005).
24. Tessier, P. M. & Lindquist, S. Prion recognition elements govern nucleation, strain specificity and species barriers. *Nature* **447,** 556–561 (2007).
25. King, C. Y. Supporting the structural basis of prion strains: induction and identification of [*PSI*] variants. *J. Mol. Biol.* **307,** 1247–1260 (2001).
26. DePace, A. H., Santoso, A., Hillner, P. & Weissman, J. S. A critical role for amino-terminal glutamine/asparagine repeats in the formation and propagation of a yeast prion. *Cell* **93,** 1241–1252 (1998).
27. Osherovich, L. Z., Cox, B. S., Tuite, M. F. & Weissman, J. S. Dissection and design of yeast prions. *PLoS Biol.* **2,** E86 (2004).
28. Crist, C. G., Nakayashiki, T., Kurahashi, H. & Nakamura, Y. [*PHI*+], a novel Sup35-prion variant propagated with non-Gln/Asn oligopeptide repeats in the absence of the chaperone protein Hsp104. *Genes Cells* **8,** 603–618 (2003).
29. Parham, S. N., Resende, C. G. & Tuite, M. F. Oligopeptide repeats in the yeast protein Sup35p stabilize intermolecular prion interactions. *EMBO J.* **20,** 2111–2119 (2001).
30. Shkundina, I. S., Kushnirov, V. V., Tuite, M. F. & Ter-Avanesyan, M. D. The role of the N-terminal oligopeptide repeats of the yeast Sup35 prion protein in propagation and transmission of prion variants. *Genetics* **172,** 827–835 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Tanaka, C. Ritter, M. Hoshino, W. Bermel and R. Riek for experimental advice; and D. Breslow, D. Cameron, S. Collins, V. Denic, K. Filaski, C. Foo, C. Gross, J. Hollien, N. Ingolia, E. Quan, E. Rodriguez and K. Tipton and other members of the Weissman laboratory for helpful discussion and critical reading of the manuscript. This research was funded by the NIH and the Howard Hughes Medical Institute.

Author Information Backbone assignments of SupNM have been deposited in the Biological Magnetic Resonance Data Bank, accession number 15379. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.S.W. (weissman@cmp.ucsf.edu).

METHODS

SupNM isotope labelling. Uniformly ^{15}N labelled SupNM was overexpressed by growing 3 l of the bacterial strain BL21/DE3 harbouring a plasmid encoding SupNM-His6× under control of the T7 promoter, to a D_{600} value of 0.3 in Luria–Bertani medium at 37 °C as described[31]. Cells were spun down, washed once in 10 ml of ^{15}N-M9 medium, and resuspended in 1 l of ^{15}N-M9 medium. After 30 min at 37 °C, expression was induced by the addition of 400 μM isopropyl β-D-thiogalactoside and after 4 h the cells were harvested. Uniformly labelled ^{13}C,^{15}N-SupNM was overexpressed in the same manner, using 1 g of ^{13}C-glucose per litre of M9 medium. ^{15}N-Lys-SupNM protein was overexpressed as described above with medium supplemented with ^{15}N-Lys. Specifically labelled ^{15}N-Glx-, Asn/Asp (Asx)-, Tyr-, Phe-, Leu- and ^{13}C,^{15}N-Glx,Asx/Tyr-SupNM proteins were overexpressed as described above by using the bacterial strain DL39 with M9 medium supplemented with the indicated ^{15}N amino acid (Asp and Glu for Asx and Glx, respectively) as well as all other amino acids unlabelled as described[32]. All ^{15}N-labelled and ^{13}C-labelled amino acids were purchased from Isotec. Protein purification was performed as described[26].

H/D exchange. ^{15}N-SupNM seeded fibres of each strain were made as described[5] and spun down at 150,000*g* for 1 h and resuspended in the equivalent buffer in D_2O at pH 2.5. Fibres were then passed through a 22-gauge needle to generate short fibres of relatively uniform size (about 100–300 nm) as assayed by AFM. To start the exchange, the pH was adjusted to 7.0. After the desired time, the pH was adjusted back to 2.5 and fibres were centrifuged at 126,000*g* for 25 min. The pellet was washed once with 5 mM DCl in D_2O and then centrifuged again at 126,000*g* for 20 min. The pellet was then frozen, freeze-dried, and stored at −80 °C until NMR acquisition. All steps, including H/D exchange, were performed at 4 °C. Fibre conformation purity was assayed by infecting samples from before and after the H/D exchange time course into [*psi*$^-$] and the conformation was determined by 'weak' or 'strong' colour phenotype as described[5]. All preparations were of greater than 95% purity as judged by this method. The NMR spectra were processed with nmrPipe[33], and peak integrations were performed with Sparky[34]. Each ^{15}N-HSQC spectrum for the time course was normalized for protein concentration by using integrations of the unexchangeable side-chain peaks in comparison with the internal DSS standard in the one-dimensional control spectra. Exchange of fibres occurred mainly at neutral pH, because when the pH was left at 2.5, peak intensities decreased slightly, which is consistent with predicted intrinsic rates at 4 °C and pH 2.5, and exchanged to a much fuller extent after the pH had been shifted to 7.0 (data not shown). Estimated minimum peak intensity was calculated by averaging the baseline for the fitted curves of the fastest-exchanging residues. Sequence-specific assignments of the backbone Hn and N resonances were achieved with triple-resonance three-dimensional HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, HN(CA)NH and CT-HNCA[35], on uniformly labelled ^{13}C,^{15}N-SupNM. HNCO, CT-HNCA, HN(CA)NH and HN(COCA)NH experiments were also performed on SupNM specifically labelled with ^{13}C,^{15}N-Glx and ^{13}C,^{15}N-Asx/Tyr. HNCO, CT-HNCA and HN(COCA)NH experiments used semi-constant time ^{15}N evolution to increase resolution[36]. Three-dimensional ^{15}N-separated HSQC-nuclear Overhauser enhancement spectroscopy (t_m = 400 ms) and HSQC-total correlation spectroscopy (t_m = 100 ms) experiments were also performed on uniformly labelled ^{15}N-SupNM and on specifically labelled ^{15}N-Asx, Lys, Glx and Tyr-SupNM. All spectra were recorded on Bruker Avance800 or DRX500M spectrometers equipped with cryoprobes with actively shielded *Z* gradients at 298 K. Assignments were performed with the program ansig[37].

31. Santoso, A., Chien, P., Osherovich, L. Z. & Weissman, J. S. Molecular basis of a yeast prion species barrier. *Cell* **100,** 277–288 (2000).
32. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **177,** 44–73 (1989).
33. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
34. Goddard, T. D. & Kneller, D. G. SPARKY 3.112, University of California, San Francisco (2006).
35. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34,** 93–158 (1999).
36. Sun, Z. Y., Frueh, D. P., Selenko, P., Hoch, J. C. & Wagner, G. Fast assignment of ^{15}N-HSQC peaks using high-resolution 3D HNcocaNH experiments with non-uniform sampling. *J. Biomol. NMR* **33,** 43–50 (2005).
37. Kraulis, P. J. ANSIG: A program for the assignment of protein ^{1}H ^{2}D NMR spectra by interactive graphics. *J. Magn. Reson.* **84,** 627–633 (1989).

Vol 449|13 September 2007|doi:10.1038/nature06115

LETTERS

Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU

Mark J. Kiel[1], Shenghui He[1], Rina Ashkenazi[2], Sara N. Gentry[2], Monica Teta[3], Jake A. Kushner[3], Trachette L. Jackson[2] & Sean J. Morrison[1]

Stem cells are proposed to segregate chromosomes asymmetrically during self-renewing divisions so that older ('immortal') DNA strands are retained in daughter stem cells whereas newly synthesized strands segregate to differentiating cells[1–6]. Stem cells are also proposed to retain DNA labels, such as 5-bromo-2-deoxyuridine (BrdU), either because they segregate chromosomes asymmetrically or because they divide slowly[5,7–9]. However, the purity of stem cells among BrdU-label-retaining cells has not been documented in any tissue, and the 'immortal strand hypothesis' has not been tested in a system with definitive stem cell markers. Here we tested these hypotheses in haematopoietic stem cells (HSCs), which can be highly purified using well characterized markers. We administered BrdU to newborn mice, mice treated with cyclophosphamide and granulocyte colony-stimulating factor, and normal adult mice for 4 to 10 days, followed by 70 days without BrdU. In each case, less than 6% of HSCs retained BrdU and less than 0.5% of all BrdU-retaining haematopoietic cells were HSCs, revealing that BrdU has poor specificity and poor sensitivity as an HSC marker. Sequential administration of 5-chloro-2-deoxyuridine and 5-iodo-2-deoxyuridine indicated that all HSCs segregate their chromosomes randomly. Division of individual HSCs in culture revealed no asymmetric segregation of the label. Thus, HSCs cannot be identified on the basis of BrdU-label retention and do not retain older DNA strands during division, indicating that these are not general properties of stem cells.

The immortal strand hypothesis was proposed as a mechanism by which stem cells could avoid accumulating mutations that arise during DNA replication[2]. Whereas most cells segregate their chromosomes randomly[1,10], it was argued that adult stem cells in steady-state tissues might retain older DNA strands during asymmetric self-renewing divisions, segregating newly synthesized strands to daughter cells fated to differentiate (Fig. 1a). Evidence has supported this model in some epithelial stem cells[1], neural stem cells[3], mammary epithelial progenitors[4] and muscle satellite cells[5,6]. A related idea is that adult stem cells in steady-state tissues might consistently retain DNA labels. This could be because chromosomes segregate randomly but stem cells divide more infrequently than other cells (Fig. 1b), or alternatively because the older DNA strand is labelled and segregated asymmetrically (Fig. 1c). Tritiated thymidine[8] or histone[7] label-retaining cells from the hair follicle are enriched for epithelial stem cells, although the purity remains uncertain. Label-retaining cells have also been identified in the haematopoietic system[9,11], in mammary epithelium[12], in intestinal epithelium[1,13] and in the heart[14], but the purity of stem cells among these label-retaining cells has not been tested. As a result, it remains unclear whether label retention can consistently identify stem cells with specificity or sensitivity.

Under steady-state conditions in adult bone marrow, all HSCs divide regularly but infrequently[15] to sustain haematopoiesis and to maintain nearly constant numbers of HSCs. As a result of this observation, as well as the finding that HSC divisions yield asymmetric outcomes in culture[16], it has been proposed that adult HSCs divide asymmetrically[16], although the rarity of HSCs *in vivo* and their

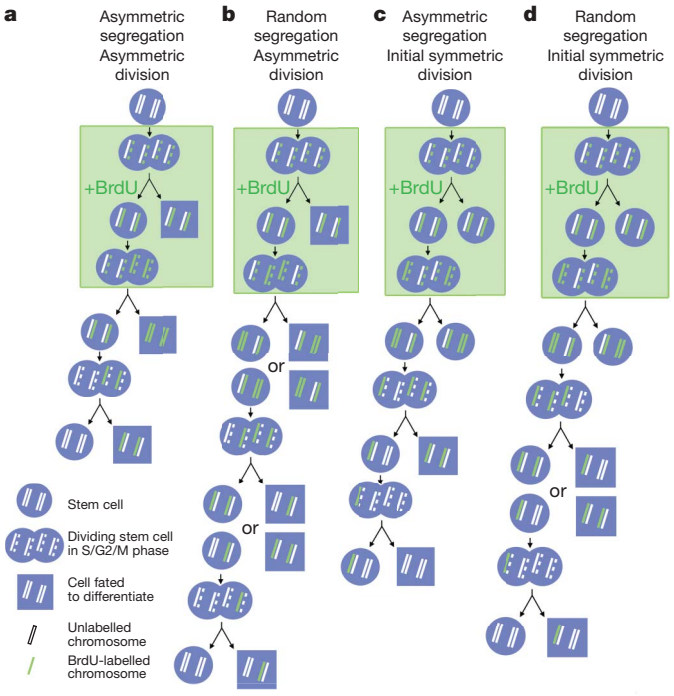


Figure 1 | Contrasting predictions regarding stem cell labelling on the basis of the immortal strand model versus random chromosome segregation. **a**, According to the immortal strand model[2], stem cells divide asymmetrically under steady-state conditions and BrdU is incorporated into newly synthesized DNA strands that are asymmetrically segregated into differentiating daughter cells with each round of division, such that stem cells retain only the unlabelled older DNA strands. **b**, In contrast, if chromosomes segregate randomly, then BrdU-labelled chromosomes will be stochastically lost over multiple rounds of divisions. **c**, In the immortal strand model, if stem cells divide symmetrically then BrdU can be incorporated into DNA strands that become the 'older' strands once stem cells resume asymmetric division. Under these circumstances, the $BrdU^+$ older strands would be retained indefinitely in stem cells. **d**, In contrast, if chromosome segregation is random then $BrdU^+$ chromosomes are stochastically lost over time after BrdU is discontinued.

[1]Howard Hughes Medical Institute, Life Sciences Institute, Department of Internal Medicine, and Centre for Stem Cell Biology, [2]Department of Mathematics, University of Michigan, Ann Arbor, Michigan 48109-2216, USA. [3]Division of Endocrinology, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA.

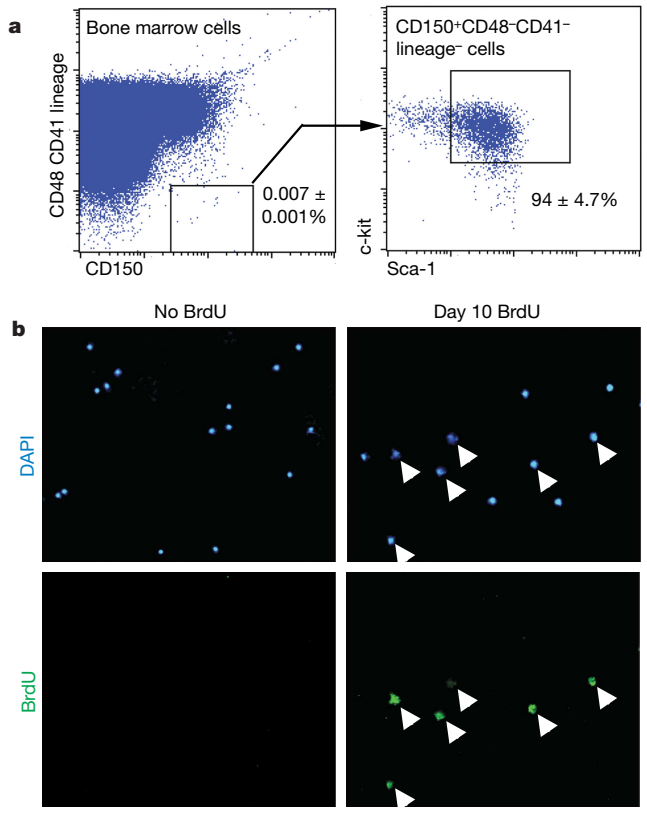
relative quiescence has made it impossible to confirm this directly. Nonetheless, if BrdU-label retention and/or asymmetric chromosome segregation are general properties of adult stem cells, then either or both of these characteristics should be evident in HSCs, depending on experimental conditions.

To test the rate at which HSCs enter the cell cycle we administered BrdU to mice for 1, 4 or 10 days, and then sorted HSCs onto microscope slides and stained with an anti-BrdU antibody. HSCs were sorted as $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ cells (Fig. 2a). This population contains all of the detectable HSC activity in bone marrow, and 47% of single cells from this population give long-term multilineage reconstitution in irradiated mice[17]. After 1–10 days of BrdU treatment, 51–94% of whole bone marrow cells and 6.5–45.9% of $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs became $BrdU^+$ (Fig. 2c, d). We calculated the rate at which HSCs entered the cell cycle[15] to be 6.0% per day (Fig. 2e). Consistent with this, only 3.2% of $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ cells were in S/G2/M phase of the cell cycle at any one time (Supplementary Fig. 3). These results are similar to a previous study that identified HSCs using different markers[15].

The linearity of BrdU incorporation over time (Fig. 2e) suggests that most HSCs divide at a similar rate. If a minority of HSCs divided more rapidly, consistently more than 6.0% of HSCs should have incorporated BrdU after one day; however, we did not observe this (Fig. 2d). If a minority of HSCs were more deeply quiescent than most other HSCs, these HSCs should remain BrdU-negative, even after long periods of BrdU treatment. This also has not been observed, because more than 99% of HSCs are labelled after 6 months of BrdU treatment[15]. Therefore, there is no evidence for more rapidly dividing or more slowly dividing subsets of long-term self-renewing HSCs under steady-state conditions, although we cannot exclude the possibility that a minority of HSCs divide more slowly than 6.0% per day.

To evaluate retention of the BrdU label, we administered BrdU for 10 days, followed by 70 days without BrdU (a 70-day 'chase'), like previous studies in the haematopoietic system[9,11]. Given that 6.0% of HSCs entered the cell cycle each day and 46% of HSCs were labelled after 10 days of BrdU treatment (Fig. 2d), we modelled the fraction of HSCs that would be expected to retain BrdU over time (Fig. 3a; see Methods for explanation).

Under these experimental conditions, if HSCs follow the immortal strand model they should lose their BrdU label one division after BrdU is discontinued because the labelled chromosomes would be segregated to differentiating daughter cells (Fig. 1a); therefore, only 0.6% of HSCs would be expected to retain BrdU after the 70-day chase (Fig. 3a). If HSCs segregate chromosomes randomly, then BrdU would be lost stochastically over time and the fraction of $BrdU^+$ HSCs after a 70-day chase would depend on the threshold of BrdU required for detection. If the threshold is equivalent to $0.5n$ labelled chromosomes (one-quarter of the genome), this level of BrdU dilution could be achieved in cells that had divided 1 or 2 times after BrdU was discontinued (depending on whether they had divided once or twice in the presence of BrdU), and only 1.4% of HSCs would be expected to retain BrdU after the 70-day chase (Fig. 3a). In contrast, if the threshold of detection is equivalent to $0.0625n$ labelled chromosomes, this could be achieved on average in cells that had divided 4 or 5 times after BrdU was discontinued, and 19.8% of HSCs would be expected to retain BrdU after the 70-day



c **$BrdU^+$ bone marrow cells (%)**

Experiment	Day 0	Day 1	Day 4	Day 10
1	0.0	49.6	78.3	92.0
2	0.0	49.3	91.0	96.5
3	0.0	53.6		93.3
Mean ± s.d.	0.0 ± 0.0	51 ± 2.4	85	94 ± 2.3

d **$BrdU^+$ HSCs (%)**

Experiment	Day 0	Day 1	Day 4	Day 10
1	0.0	4.5	25.0	44.6
2	0.0	5.6	23.6	43.2
3	0.0	5.0		50.0
4	0.0	10.9		
Mean ± s.d.	0.0 ± 0.0	6.5 ± 3.0	24	46 ± 3.6

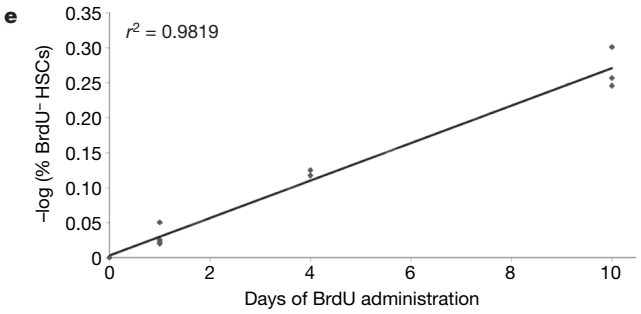


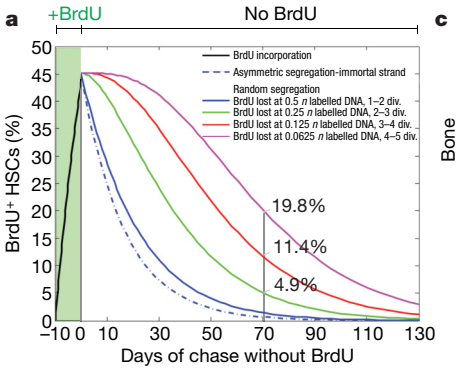
Figure 2 | Six per cent of HSCs stochastically enter the cell cycle each day. **a**, HSCs can be isolated by flow cytometry as $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ cells; these represent only 0.0066 ± 0.0003% (0.007% × 94%) of bone marrow cells but contain all detectable HSC activity and are very highly enriched for HSCs (nearly 50% of single cells give long-term multilineage reconstitution in irradiated mice[17]). **b**, BrdU incorporation into HSCs (arrowheads) is evaluated by immunofluorescence after sorting HSCs onto microscope slides (DAPI is a nuclear stain). **c**, **d**, The percentage of $BrdU^+$ bone marrow cells (**c**) and $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs (**d**) after various periods of BrdU administration (3–4 independent experiments with 3–4 mice per experiment and 200–400 bone marrow cells or 100–400 HSCs counted, respectively, per mouse). Standard deviations are shown for means that are based on at least three independent experiments. **e**, The percentage of HSCs that enter the cell cycle each day (6%) can be derived by plotting the negative logarithm of the percentage of HSCs that were $BrdU^-$ over time[15].

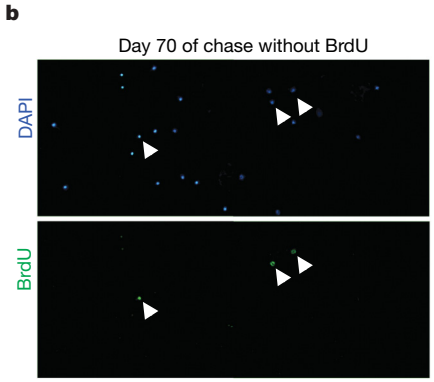
chase (Fig. 3a). Thus, these calculations predict that few (<20%) HSCs should retain BrdU after a 70-day chase, irrespective of how chromosomes are segregated.

To test this we administered BrdU for 10 days to adult mice and then stained whole bone marrow cells and $CD150^+CD48^-CD41^-$ $lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs after 40, 70 and 120 days of chase. After 70 days of chase, 0.4 ± 0.1% (mean ± s.d.) of bone marrow cells and 4.6 ± 0.9% of HSCs were $BrdU^+$ (Fig. 3c). This demonstrates that, as predicted (Fig. 3a), few HSCs retain BrdU. Moreover, only 2.0 ± 1.0% of HSCs were $BrdU^+$ after 120 days of chase, demonstrating that the frequency of BrdU-retaining HSCs continues to decline over time rather than there being a deeply quiescent subset of HSCs that retains BrdU indefinitely. Although BrdU-label-retaining cells were enriched in HSCs by tenfold compared to whole bone marrow cells, the rarity of HSCs means that only 0.08% of $BrdU^+$ bone marrow cells were HSCs (0.0066% × 4.6%/0.4%). BrdU-label retention is therefore a very insensitive and nonspecific marker of HSCs because most HSCs did not retain detectable BrdU, and only rare BrdU-label-retaining cells were HSCs. Very similar results were obtained when we used flow cytometry to detect BrdU incorporation (Supplementary Fig. 1) or when HSCs were isolated using different surface markers ($c\text{-}kit^+Flk\text{-}2^-lineage^-Sca\text{-}1^+$ cells; Supplementary Fig. 2). These data are most consistent with random chromosome segregation by HSCs and the failure to detect BrdU after approximately three divisions in the absence of BrdU (Fig. 3a).

According to the immortal strand model, stem cells can incorporate BrdU into DNA strands that become the 'older' strands during symmetric cell divisions, and these labelled DNA strands would be retained indefinitely by stem cells that resume asymmetric divisions (Fig. 1c). To test this, we administered BrdU to newborn mice for 10 days or to cyclophosphamide and granulocyte colony-stimulating factor (cyclophosphamide/G-CSF)-treated mice for 4 days. The absolute number of HSCs expands markedly in both newborn[18] and cyclophosphamide/G-CSF-mobilized mice[19] (indicating symmetric divisions), before stabilizing to steady-state levels as mice enter adulthood or after G-CSF is discontinued. After 10 days of BrdU treatment in neonatal mice, 93 ± 3.7% of bone marrow cells and 80 ± 11% of $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs were $BrdU^+$. Seventy days later, 0.1% of bone marrow cells and 6.4 ± 2.5% of $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs were $BrdU^+$ (Fig. 3d). After 4 days of BrdU treatment in cyclophosphamide/G-CSF mobilized mice, 94 ± 3% of $CD150^+CD48^-CD41^-$ $lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs were $BrdU^+$. Seventy days later, 0.1% of bone marrow cells and 2.1 ± 1.1% of $CD150^+CD48^-CD41^-lineage^-$ $Sca\text{-}1^+c\text{-}kit^+$ HSCs were $BrdU^+$ (Fig. 3d). Thus, even when BrdU was administered to symmetrically dividing HSCs, only 2–6% of HSCs retained the label and only 0.2–0.4% of BrdU-retaining bone marrow cells were HSCs. We were unable to identify any context in which BrdU-label retention identified HSCs with sensitivity or specificity, and none of these results was consistent with the immortal strand hypothesis.

To address the possibility that the HSCs in the above experiments might have continued to divide symmetrically after BrdU was discontinued, we also administered BrdU to mice from 20 to 29 days postnatally. HSCs are thought to transition from rapidly dividing cells that have a fetal phenotype to relatively quiescent cells that have an adult phenotype between 21 and 28 days postnatally[18]. We obtained similar results, with only 6.5 ± 1.1% of HSCs retaining





c **BrdU⁺ cells from adult mice (%)**

	Experiment	40 days chase	70 days chase	120 days chase
Bone marrow cells	1	0.3	0.4	0.3
	2	0.6	0.3	0.4
	3		0.5	0.1
	Mean ± s.d.	0.4	0.4 ± 0.1	0.3 ± 0.1
	Experiment	40 days chase	70 days chase	120 days chase
HSCs	1	7.4	4.0	1.5
	2	14.5	5.7	1.4
	3	8.3	3.4	3.1
	4	6.9	5.0	
	5		4.8	
	Mean ± s.d.	9.3 ± 3.5	4.6 ± 0.9	2.0 ± 1.0

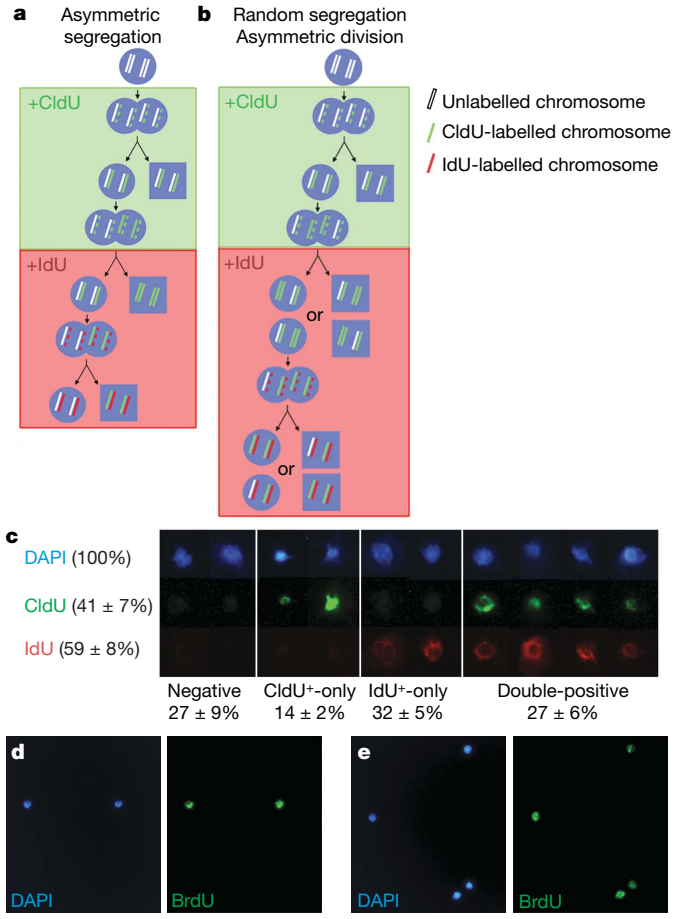
d **BrdU⁺ cells from neonatal or mobilized mice (%)**

	Experiment	70 days chase neonate	120 days chase neonate	70 days chase mobilized
Bone marrow cells	1	0.0	0.1	0.0
	2	0.2	0.1	0.0
	3	0.2	0.0	0.2
	Mean ± s.d.	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	Experiment	70 days chase neonate	120 days chase neonate	70 days chase mobilized
HSCs	1	6.1	2.5	2.8
	2	5.6	0.8	2.6
	3	4.3	2.1	0.9
	4	10.8		
	5	5.3		
	Mean ± s.d.	6.4 ± 2.5	1.8 ± 0.9	2.1 ± 1.1

Figure 3 | Few HSCs retain BrdU, and most BrdU-retaining bone marrow cells are not HSCs. **a**, A model is shown that predicts the fraction of HSCs that retain BrdU over time after administering BrdU for 10 days, depending on whether chromosomes segregate asymmetrically or randomly and on the threshold of BrdU that can be detected by immunofluorescence (0.5*n*, 0.25*n*, 0.125*n* or 0.0625*n* labelled DNA). **b**, $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+$ $c\text{-}kit^+$ HSCs were sorted onto a microscope slide after 10 days BrdU administration and 70 days chase (without BrdU). Arrowheads identify $BrdU^+$ cells. **c**, Shown is the frequency of $BrdU^+$ bone marrow cells and HSCs after 10 days of BrdU administration and 40, 70 or 120 days of chase. Standard deviations are shown for means that are based on at least three independent experiments. **d**, Shown is the frequency of $BrdU^+$ bone marrow cells and HSCs after 10 days BrdU administration to neonatal mice followed by 70 or 120 days of chase, or after 4 days BrdU administration to cyclophosphamide/G-CSF-mobilized mice followed by 70 days of chase. All data are based on 3–5 independent experiments with 3 mice per experiment and 400–700 bone marrow cells or 300–400 HSCs counted per mouse.

BrdU after a 70-day chase. There was, therefore, no period during neonatal development when BrdU could be administered in a way that resulted in retention of BrdU within significant numbers of HSCs.

To test the immortal strand model directly, we treated mice with 5-chloro-2-deoxyuridine (CldU) for 10 days and then with 5-iodo-2-deoxyuridine (IdU) for 10 days. If HSCs segregate older and younger DNA strands asymmetrically, then HSCs should rarely incorporate both CldU and IdU under steady-state conditions because newly synthesized (labelled) DNA strands should be segregated to differentiating daughter cells after each division (Fig. 4a). In contrast, if HSCs segregate older and younger DNA strands randomly then CldU-labelled HSCs should have the same chance of incorporating IdU as unlabelled cells, and approximately 25% (50% × 50%) of HSCs should be double-labelled (Fig. 4b).



f **Single cell culture of BrdU-labeled HSCs**

Number of cells harvested from individual colony (>1 cell)	Colonies with only $BrdU^-$ cells	Colonies with only $BrdU^+$ cells	Colonies with $BrdU^+$ and $BrdU^-$ cells
2 cells	123	86	0
3 cells	38	50	0
4 cells	23	26	0
Total	184	162	0

Figure 4 | HSCs segregate chromosomes randomly *in vivo* and *in vitro*. **a**, By the immortal strand model, stem cells sequentially exposed to CldU (for 10 days) and then IdU (for 10 days) would not incorporate both labels, with the exception of rare cells in the S/G2/M phase of their first division after switching from CldU into IdU (expected frequency <3%). **b**, In contrast, if chromosome segregation is random then $CldU^+$ stem cells would have the same probability of incorporating IdU as unlabelled cells (expected frequency of $CldU^+IdU^+$ HSCs is ~25%). **c**, CldU was administered to mice for 10 days followed by IdU for 10–11 days, and $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+$ $c\text{-}kit^+$ HSCs were stained. Examples of HSCs that incorporated neither label, only CldU, only IdU or both labels are shown (mean ± sd; data are based on two independent experiments with 2 or 3 mice per experiment and 100–250 HSCs per mouse). HSCs from BrdU-treated mice divided once (**d**) or twice (**e**) in culture to form daughter cells. **f**, The progeny of these HSCs were either all $BrdU^+$ or all $BrdU^-$. We detected no clones in which label was asymmetrically segregated to a subset of daughter cells.

After 10 days of CldU followed by 10–11 days of IdU, we observed that 14% of HSCs incorporated only CldU, 32% incorporated only IdU and 27% incorporated both CldU and IdU (Fig. 4c and Supplementary Fig. 4). The frequency of $CldU^+IdU^+$ cells (27%) was therefore similar to the product of the frequencies of total $CldU^+$ cells and total IdU^+ cells (41% × 59% = 24%), indicating that $CldU^+$ cells had a similar probability of incorporating IdU as the other cells. We repeated this experiment by administering mice with CldU for 60 days followed by IdU for 15 days, and found the frequency of $CldU^+IdU^+$ cells (63%) was again similar to the product of the frequencies of total $CldU^+$ cells and total IdU^+ cells (73% × 84% = 61%). These results were not significantly affected by a slow clearance of CldU from mice, because CldU was cleared in less than 1 day after being discontinued from the drinking water (Supplementary Fig. 5). These observations directly contradict a key prediction made by the immortal strand hypothesis, but are as would be expected by random chromosome segregation.

The foregoing experiments left the formal possibility that if HSCs divide by a combination of symmetric and asymmetric divisions *in vivo* we might underestimate the frequency of HSCs that retain older DNA strands. To address this we examined the division of individual HSCs in culture that were isolated from mice treated for 10 days with BrdU. Single $CD150^+CD48^-CD41^-lineage^-Sca\text{-}1^+c\text{-}kit^+$ HSCs were sorted into cultures under conditions in which half of HSC divisions give asymmetric outcomes (daughter cells with different developmental potentials)[16]. After 2–3 days of culture we observed a total of 346 colonies in which HSCs had divided once (2 daughter cells) or twice (3 or 4 daughter cells). Either all daughter cells were $BrdU^+$ (162 colonies, 46%) or all daughter cells were $BrdU^-$ (184 colonies; Fig. 4f), as would be expected by random chromosome segregation given that 46% of HSCs incorporate BrdU *in vivo* after 10 days (Fig. 2d). We observed no colonies containing a mixture of $BrdU^+$ and $BrdU^-$ cells after one or two rounds of division. Thus, these *in vitro* experiments on individual HSCs failed to detect any asymmetric segregation of labelled chromosomes.

Our results were not confounded by the effects of BrdU, CldU or IdU on HSC proliferation, survival or DNA repair. The cell cycle status (Supplementary Fig. 3) and frequency (data not shown) of HSCs was not affected by these treatments. BrdU, CldU and IdU incorporation from DNA repair was not detectable (Supplementary Fig. 3).

Our results indicate that BrdU-label retention is neither a sensitive nor a specific marker of HSCs. Nonetheless BrdU-label retention could be a better marker of stem cells in other tissues. Moreover, histone–green fluorescent protein retention[7] may do a better job of marking stem cells, including HSCs, because it can be selectively expressed in subsets of progenitors and may be retained with different kinetics than BrdU. Our data demonstrate the need to test the sensitivity and specificity of BrdU and other label-retention markers before assuming they mark stem cells with fidelity.

Our data also demonstrate that the immortal strand model[2] does not apply to HSCs and cannot be considered a general model of stem cell division. Our data do not address whether stem cells from other tissues asymmetrically segregate chromosomes or whether HSCs segregate a limited number of chromosomes asymmetrically[10]. Nonetheless, asymmetric chromosome segregation cannot be a mechanism by which HSCs avoid accumulating mutations over time.

METHODS SUMMARY

BrdU, CldU and IdU administration. All experiments used C57BL/Ka-CD45.2:Thy-1.1 mice. For experiments in adult mice, BrdU (Sigma) was

administered when the mice were 8–10 weeks of age. Mice were given an intraperitoneal injection of 100 mg BrdU per kg body mass in Dulbecco's phosphate buffered saline (DPBS; Gibco) and were maintained on 1 mg ml^{-1} BrdU in the drinking water for 1–10 days. Amber bottles containing BrdU water were changed every 1–3 days. For retention studies, BrdU water was replaced with regular water and the mice were maintained for 40–120 days before analysis.

BrdU injections into neonatal mice were performed as described[20]. Beginning within 3 days after birth, neonatal mice were injected subcutaneously with 50 mg BrdU per kg body weight twice daily for 10 days. Mice were weighed every 2 days and the dose of BrdU was adjusted.

To assess BrdU retention after cytokine mobilization, adult mice were injected with cyclophosphamide (200 mg kg^{-1}; Bristol-Myers Squibb) and then on each of the 4 subsequent days they were injected with 250 μg kg^{-1}day^{-1} of human G-CSF (Amgen)[19]. On the fourth day of G-CSF injection, a single intraperitoneal injection of 100 mg BrdU per kg body mass was given and the mice were put on 1 mg ml^{-1} BrdU water for 4 additional days. The mice were then returned to regular water for 70 days before analysis.

For CldU and IdU experiments, mice were given an intraperitoneal injection of 100 mg CldU per kg body mass in DPBS and were maintained on drinking water containing 1 mg ml^{-1} CldU (Sigma) for 10 days. Mice were then given an intraperitoneal injection of 100 mg IdU per kg body mass in DPBS, and were switched to drinking water containing 1 mg ml^{-1} IdU (Sigma) for 10–11 days before being killed.

For details related to the flow cytometric isolation of HSCs and CldU, IdU and BrdU staining, see Methods.

BrdU segregation in cultured HSCs. Single CD150$^+$CD48$^-$CD41$^-$lineage$^-$Sca-1$^+$ c-kit$^+$ HSCs were sorted from BrdU-treated mice into a V-bottom 96-well plate containing Stempro-34 medium (Invitrogen) supplemented with 2 mM L-glutamine, 50 μM 2-mercaptoethanol (Sigma), murine IL-3 (10 ng ml^{-1}), murine SCF (100 ng ml^{-1}) and murine Tpo (100 ng ml^{-1}; all cytokines were obtained from R&D Systems) with 10% charcoal absorbed fetal bovine serum (Cocalico Biologicals Inc.), and were cultured for 2–3 days in low-oxygen chambers[21]. For analysis, plates were centrifuged at 500*g* for 10 min; cells from each colony were then pipetted onto individual wells of Teflon-printed glass slides and were allowed to dry overnight before staining for BrdU and 4,6-diamidino-2-phenylindole (DAPI).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 June; accepted 25 July 2007.
Published online 29 August 2007.

1. Potten, C. S., Hume, W. J., Reid, P. & Cairns, J. The segregation of DNA in epithelial stem cells. *Cell* **15,** 899–906 (1978).
2. Cairns, J. Mutation selection and the natural history of cancer. *Nature* **255,** 197–200 (1975).
3. Karpowicz, P. *et al.* Support for the immortal strand hypothesis: neural stem cells partition DNA asymmetrically *in vitro*. *J. Cell Biol.* **170,** 721–732 (2005).
4. Smith, G. H. Label-retaining epithelial cells in mouse mammary gland divide asymmetrically and retain their template DNA strands. *Development* **132,** 681–687 (2005).
5. Shinin, V., Gayraud-Morel, B., Gomes, D. & Tajbakhsh, S. Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. *Nature Cell Biol.* **8,** 677–687 (2006).
6. Conboy, M. J., Karasov, A. O. & Rando, T. A. High incidence of non-random template strand segregation and asymmetric fate determination in dividing stem cells and their progeny. *PLoS Biol.* **5,** e102 (2007).
7. Tumbar, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303,** 359–363 (2004).
8. Cotsarelis, G., Sun, T. T. & Lavker, R. M. Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell* **61,** 1329–1337 (1990).
9. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425,** 836–841 (2003).
10. Armakolas, A. & Klar, A. J. Cell type regulates selective segregation of mouse chromosome 7 DNA strands in mitosis. *Science* **311,** 1146–1149 (2006).
11. Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118,** 149–161 (2004).
12. Welm, B. E. *et al.* Sca-1$^+$ cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev. Biol.* **245,** 42–56 (2002).
13. Potten, C. S., Owen, G. & Booth, D. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* **115,** 2381–2388 (2002).
14. Urbanek, K. *et al.* Stem cell niches in the adult mouse heart. *Proc. Natl Acad. Sci. USA* **103,** 9226–9231 (2006).
15. Cheshier, S. H., Morrison, S. J., Liao, X. & Weissman, I. L. *In vivo* proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **96,** 3120–3125 (1999).
16. Takano, H., Ema, H., Sudo, K. & Nakauchi, H. Asymmetric division and lineage commitment at the level of hematopoietic stem cells: inference from differentiation in daughter cell and granddaughter cell pairs. *J. Exp. Med.* **199,** 295–302 (2004).
17. Kiel, M. J., Yilmaz, O. H., Iwashita, T., Terhorst, C. & Morrison, S. J. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121,** 1109–1121 (2005).
18. Bowie, M. B. *et al.* Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *J. Clin. Invest.* **116,** 2808–2816 (2006).
19. Morrison, S. J., Wright, D. & Weissman, I. L. Cyclophosphamide/granulocyte colony-stimulating factor induces hematopoietic stem cells to proliferate prior to mobilization. *Proc. Natl Acad. Sci. USA* **94,** 1908–1913 (1997).
20. Taylor, G., Lehrer, M. S., Jensen, P. J., Sun, T. T. & Lavker, R. M. Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell* **102,** 451–461 (2000).
21. Morrison, S. J. *et al.* Culture in reduced levels of oxygen promotes clonogenic sympathoadrenal differentiation by isolated neural crest stem cells. *J. Neurosci.* **20,** 7370–7376 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Howard Hughes Medical Institute, the National Institute on Aging (NIH), and the US Army Research Laboratory/Office. Flow cytometry was partially supported by the UM-Comprehensive Cancer Centre and the UM-Multipurpose Arthritis Centre. Antibody production was partially supported by the Rheumatic Core Disease Centre. M.J.K. was supported by a University of Michigan Cancer Biology Training Grant. The authors thank D. Adams and M. White for flow cytometry and E. Smith (Hybridoma Core Facility) for antibody production.

Author Contributions M.J.K. performed all experiments and interpreted results. S.H. assisted in the design and interpretation of many experiments and helped to perform some experiments. R.A., S.N.G. and T.L.J. generated the mathematical model of BrdU retention over time (Fig. 3a). M.T. and J.A.K. developed the protocol for double-labelling cells with CldU and IdU. S.J.M. participated in the design and interpretation of experiments, and wrote the paper with M.J.K. and S.H.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to S.J.M. (seanjm@umich.edu).

METHODS

Flow cytometric isolation of HSCs. Bone marrow cells were flushed from the long bones (tibias and femurs) with Hank's buffered salt solution (HBSS) without calcium or magnesium supplemented with 2% heat-inactivated calf serum (GIBCO). Cells were triturated and filtered through a nylon screen (45 μm, Sefar America) to obtain a single-cell suspension. For isolation of Flk2$^-$lineage$^-$Sca-1$^+$ c-kit$^+$ HSCs, whole bone marrow cells were incubated with phycoerythrin-conjugated monoclonal antibodies against lineage markers including B220 (6B2), CD3 (KT31.1), CD4 (GK1.5), CD8 (53-6.7), Gr-1 (8C5), Mac-1 (M1/70), Flk-2 (A2F10.1), Ter119 and IgM in addition to fluorescein isothiocyanate (FITC)-conjugated anti-Sca-1 (Ly6A/E) and biotin-conjugated anti-c-kit (2B8) antibodies, followed by streptavidin allophycocyanin (APC)-Cy7. For isolation of CD150$^+$ CD48$^-$CD41$^-$lineage$^-$Sca-1$^+$c-kit$^+$ HSCs, whole bone marrow cells were incubated with phycoerythrin-conjugated anti-CD150 (TC15-12F12.2; BioLegend), FITC-conjugated anti-CD48 (HM48-1; BioLegend), FITC-conjugated anti-CD41 (MWReg30; BD Pharmingen), APC-conjugated anti-Sca-1 (E13-6.7) and biotin-conjugated anti-c-kit (2B8) antibodies, in addition to FITC-conjugated antibodies against Ter119, B220 (6B2), Gr-1 (8C5) and CD2 (RM2-5). HSCs were frequently pre-enriched by selecting c-kit$^+$ cells using paramagnetic microbeads and autoMACS (Miltenyi Biotec). All flow cytometry was performed on a FACSVantage SE-dual laser, three-line flow cytometer (Becton-Dickinson).

BrdU staining. Staining for BrdU on slides was performed as previously described using an anti-BrdU antibody (clone BU1/75, Accurate Chemical and Scientific Corp.)[15]. HSCs were sorted and then resorted (to ensure purity) onto glass slides in 25–100-cell spots, allowed to dry for 1 h and stored at −80 °C for up to 4 weeks. Slides were thawed at room temperature (23 °C) for 15 min, fixed in 70% ethanol at –20 °C for 30 min, rinsed in 0.1 M phosphate buffer twice, incubated in 2 M HCl with 0.8% Triton in phosphate buffer for 30 min, incubated in 0.1 M sodium borate (pH 8.5) for 15 min, and rinsed in 0.1 M phosphate buffer at room temperature, then at 37 °C, and finally at room temperature again. The slides were then incubated with 0.3% Triton in 0.1 M phosphate buffer supplemented with 5% goat serum for 1 h at 4 °C. Slides were incubated at 4 °C overnight with a primary anti-BrdU antibody that specifically recognizes BrdU and CldU but does not recognize IdU (clone BU1/75, Accurate Chemical and Scientific Corp.). Slides were then rinsed in 0.1 M phosphate buffer twice and were incubated with Alexa488-conjugated goat anti-rat IgG (Invitrogen-Molecular Probes) for 2 h at room temperature. Slides were rinsed in 0.1 M phosphate buffer twice and were incubated with DAPI for 1 h at room temperature. Finally, slides were rinsed 3 times in 0.1 M phosphate buffer, the excess buffer was shaken off and the slides were mounted in 70% glycerol in PBS. Images were gathered using an Olympus BX-51 fluorescence microscope equipped with a Cooke Pixelfly CCD camera.

In some experiments, BrdU incorporation was measured by flow cytometry using an antibody directly conjugated to APC (APC BrdU flow kit, BD Pharmingen). Sorted samples were fixed and permeabilized according to manufacturer's instructions, incubated in 2 M HCl for 30 min at room temperature, neutralized in 0.1 M sodium borate (pH 8.5), washed in 0.5% Triton in DPBS, stained with anti-BrdU APC and resuspended in DAPI (10 μg ml^{-1}) before FACS analysis. All flow cytometry was performed on a FACSVantage SE-dual laser, three-line flow cytometer (Becton-Dickinson).

CldU and IdU double-labelling. Slides were processed as for BrdU analysis with the addition of a subsequent set of staining steps using a second anti-IdU antibody (clone B44, BD Pharmingen) that does not recognize CldU. Slides were incubated at room temperature in anti-IdU antibody for 2–3 h. Immunofluorescence was developed using a Cy3-conjugated anti-mouse IgG secondary antibody with minimal cross reactivity to rat IgG (Jackson Immunoresearch). Cells isolated from mice that had only received CldU, had only received IdU, or had received neither were processed in parallel with experimental samples and demonstrated no cross-reactivity or background staining from primary or secondary antibodies.

Analysis of cell cycle distribution and cell death in HSCs. Cell cycle distribution was analysed by Hoechst 33342 (Invitrogen-Molecular Probes) staining. CD150$^+$CD48$^-$CD41$^-$lineage$^-$Sca-1$^+$c-kit$^+$ HSCs were sorted and resorted into ice-cold 70% ethanol and stored at –20 °C overnight. Cells were resuspended in PBS containing 0.02 mg ml^{-1} Hoechst 33342, incubated for 30 min and analysed by flow cytometry using an ultraviolet laser.

For activated caspase-3 staining, frozen slides bearing sorted CD150$^+$CD48$^-$ CD41$^-$lineage$^-$Sca-1$^+$c-kit$^+$ HSCs or sections through embryonic day 11 mouse forebrain were thawed at room temperature for 10 min, fixed at room temperature in 10% buffered neutral formalin (VWR) for 10 min, rinsed in 0.1 M phosphate buffer twice and blocked with 0.3% Triton in 0.1 M phosphate buffer supplemented with 5% goat serum for 1 h at 4 °C. Slides were then incubated with anti-activated caspase-3 antibody (BD Pharmingen) at room temperature for 2 h, and rinsed in 0.1 M phosphate buffer twice and incubated with Alexa488-conjugated goat anti-rabbit IgG (Invitrogen-Molecular Probes) for 1 h at room temperature. The slides were then rinsed in 0.1 M phosphate buffer twice and incubated with DAPI for 30 min at room temperature. Finally, slides were rinsed three times in 0.1 M phosphate buffer, the excess buffer was shaken off and the slides were mounted in 70% glycerol in PBS. Images were collected using a fluorescence microscope.

To test for the incorporation of BrdU due to DNA repair, BrdU was administered to adult mice for 12 h followed by irradiation with 100 rad from a Gammacell40 Extractor (MDS Nordion) followed by 48 h of further BrdU administration. CD150$^+$CD48$^-$CD41$^-$lineage$^-$Sca-1$^+$c-kit$^+$ HSCs were then sorted onto slides and stained for BrdU as described above.

Mathematical models of BrdU uptake and retention. To model the uptake and retention of BrdU in a population of stem cells we assumed that for days 0 through to T the stem cells are exposed to adequate BrdU so that cells incorporate BrdU when they divide. On the basis of our data, a random 6.0% of HSCs enter the cell cycle each day. At day T, the BrdU supply is removed and the level of BrdU incorporated into the chromosomes decreases for every cell division after day T. The rate at which BrdU is diluted from cells during this chase period depends on how the cells segregate their chromosomes. If chromosomes segregate randomly then, irrespective of whether stem cells divide asymmetrically or symmetrically with respect to daughter-cell fate, BrdU-labelled chromosomes will stochastically become diluted over time: on average, the BrdU label will be diluted by half during each round of division and multiple divisions will be required to dilute the BrdU label to the point at which it is no longer detectable by immunohistochemistry. This is modelled as case 1 (see below) for asymmetrically dividing cells. In contrast, according to the immortal strand model[2], stem cells divide asymmetrically under steady-state conditions and BrdU is preferentially incorporated into newly synthesized DNA strands that are asymmetrically segregated into differentiating daughter cells with each round of division. Under these assumptions, modelled as case 2 (see below), stem cells retain only the unlabelled older DNA strands once BrdU is withdrawn.

Case 1: random segregation of chromosomes. When chromosomes are allowed to segregate randomly, BrdU could be incorporated into one or two DNA strands within each chromosome, depending on the number of times a stem cell divides during the period of BrdU incorporation and the way in which the chromosomes segregate. To model the rate of BrdU incorporation: y_0 represents the fraction of cells without BrdU; y_1 represents the fraction of cells with one strand BrdU$^+$ after only one division; y_2 represents the fraction of cells with both strands BrdU$^+$ after two or more divisions; and α represents the proliferation rate of stem cells (we observed 6.0% of HSCs enter the cell cycle per day). Cell death was not incorporated into this model because we did not observe significant cell death or changes in HSC frequency during the experiments. The equations for uptake are as follows:

$$\frac{dy_0}{dt} = -\alpha y_0 \quad (1)$$

$$\frac{dy_1}{dt} = -\alpha y_1 + \alpha y_0 \quad (2)$$

$$\frac{dy_2}{dt} = \alpha y_1 \quad (3)$$

In equation (1), cells leave the y_0 population when they divide. In equation (2), cells from the y_0 population are added into the y_1 population through the incorporation of BrdU into one of the DNA strands. Cells leave the y_1 population when they divide. In equation (3), cells from the y_1 population are added into the y_2 population through further incorporation of BrdU through cell division.

To determine the frequency of BrdU$^+$ stem cells at any time after the addition of BrdU, we solve the system of ordinary differential equations, with all HSCs initially being unlabelled before BrdU administration. Similar equations have been used previously to model BrdU incorporation and depletion from other cells[22].

At day T (when BrdU is removed), we determine the total number of BrdU$^+$ stem cells by adding the y_1 and y_2 populations. Cells in the y_1 population have a BrdU level of 1, whereas y_2 cells have a BrdU level of up to 2. To model the rate at which these cells lose BrdU during subsequent divisions after removing BrdU, we used the following: y_{10} represents the fraction of cells, initially with one labelled DNA strand, that undergoes zero divisions after day T; y_{11} represents the fraction of cells, initially with one labelled DNA strand, that undergoes one division after day T; y_{1N} represents the fraction of cells, initially with one labelled DNA strand, that undergoes N divisions after day T; y_{20} represents the fraction of cells, initially with two labelled DNA strands, that undergoes zero divisions after day T; y_{21} represents the fraction of cells, initially with two labelled DNA strands, that

undergoes one division after day T; y_{2N} represents the fraction of cells, initially with two labelled DNA strands, that undergoes N divisions after day T. Cells move from $y_{1(k-1)}$ to y_{1k} at the proliferation rate α (cells move in y_{2k} similarly). The equations for dilution of the BrdU label are:

$$\frac{dy_{10}}{dt} = -\alpha y_{10}$$

$$\frac{dy_{11}}{dt} = -\alpha y_{11} + \alpha y_{10}$$

$$\frac{dy_{1N}}{dt} = \alpha y_{1(N-1)}$$

$$\frac{dy_{20}}{dt} = -\alpha y_{20}$$

$$\frac{dy_{21}}{dt} = -\alpha y_{21} + \alpha y_{20}$$

$$\frac{dy_{2N}}{dt} = \alpha y_{2(N-1)}$$

With each cell division, the cell's BrdU level is decreased by half. For instance, cells in y_{10} have a BrdU level of 1, cells in y_{11} have half as much BrdU on average, and cells in y_{12} have one-quarter as much BrdU on average, and so on. Similarly, cells in y_{20} have a BrdU level of 2, cells in y_{21} have BrdU level 1, cells in y_{22} have a BrdU level of 0.5, and so on. We can determine the total fraction of cells that have a BrdU level that is above a minimum detection level (which can be set at any desired level in the simulations) by adding all relevant populations at any time after day T.

This set of ordinary differential equations is solved for initial conditions that are determined by the observed results of the BrdU incorporation at time T. In this way, it is possible to plot the frequency of $BrdU^+$ HSCs over time, depending on whether it takes 1, 2, 3 or 4 rounds of division to dilute BrdU to the point at which it is no longer detectable (Fig. 3a). Note that one round of division corresponds to a 2-fold dilution of BrdU whereas four rounds of division correspond to 16-fold dilution. Our empirical data indicate that approximately three rounds of division are required to dilute BrdU to the point that it is no longer detectable in HSCs.

Case 2: asymmetric chromosome segregation. In this model, only one strand of DNA within each chromosome in stem cells can contain BrdU, irrespective of how long BrdU is administered (although the proportion of labelled stem cells will increase over time). In case 2, the equations for uptake are equation (1) (see 'case 1') and:

$$\frac{dy_1}{dt} = \alpha y_0 \quad (4)$$

In equation (4), cells from the y_0 population are added into the y_1 population through the incorporation of BrdU into one of the DNA strands.

As before, to determine the proportion of cells within each population at the end of day T, we solve the system of ordinary differential equations with initial conditions corresponding to the case in which initially all HCSs are unlabelled before BrdU administration. Cells in population y_0 do not contain any BrdU, whereas those in y_1 have a BrdU level of one.

To study the process by which cells lose BrdU label, we took all cells containing BrdU at day T (all cells in population y_1 and monitored BrdU loss through cell division in the absence of BrdU). In this case, the asymmetric segregation of chromosomes means that all dividing stem cells will lose all BrdU label in a single division in the absence of BrdU. Therefore, the fraction of stem cells that remain $BrdU^+$ at time T simplifies to the fraction of $BrdU^+$ HSCs that do not divide after removing BrdU. Cells move from y_{10} (and lose their BrdU label) at the proliferation rate α. The equation for loss of BrdU label is:

$$\frac{dy_{10}}{dt} = -\alpha y_{10}$$

This differential equation with initial condition $y_{10}(0) = y_1(T)$ determines the fraction of HSCs that retain BrdU over time, according to the immortal strand hypothesis.

22. Bonhoeffer, S., Mohri, H., Ho, D. & Perelson, A. S. Quantification of cell turnover kinetics using 5-bromo-2′-deoxyuridine. *J Immunol.* **164**, 5049–5054 (2000).

LETTERS

The structural basis for activation of plant immunity by bacterial effector protein AvrPto

Weiman Xing[1,2], Yan Zou[1,3]*, Qun Liu[4]*, Jianing Liu[1], Xi Luo[1], Qingqiu Huang[3], She Chen[1], Lihuang Zhu[3], Ruchang Bi[2], Quan Hao[4], Jia-Wei Wu[5], Jian-Min Zhou[1] & Jijie Chai[1]

Pathogenic microbes use effectors to enhance susceptibility in host plants. However, plants have evolved a sophisticated immune system to detect these effectors using cognate disease resistance proteins[1], a recognition that is highly specific, often elicits rapid and localized cell death, known as a hypersensitive response, and thus potentially limits pathogen growth[2–5]. Despite numerous genetic and biochemical studies on the interactions between pathogen effector proteins and plant resistance proteins, the structural bases for such interactions remain elusive. The direct interaction between the tomato protein kinase Pto and the *Pseudomonas syringae* effector protein AvrPto is known to trigger disease resistance and programmed cell death[6,7] through the nucleotide-binding site/leucine-rich repeat (NBS-LRR) class of disease resistance protein Prf[8]. Here we present the crystal structure of an AvrPto–Pto complex. Contrary to the widely held hypothesis that AvrPto activates Pto kinase activity, our structural and biochemical analyses demonstrated that AvrPto is an inhibitor of Pto kinase *in vitro*. The AvrPto–Pto interaction is mediated by the phosphorylation-stabilized P+1 loop and a second loop in Pto, both of which negatively regulate the Prf-mediated defences in the absence of AvrPto in tomato plants. Together, our results show that AvrPto derepresses host defences by interacting with the two defence-inhibition loops of Pto.

The *Pseudomonas syringae* pv. *tomato* effector proteins AvrPto and AvrPtoB—on injection into host cells by the type III secretion system[9–11]—promote virulence in susceptible plants, but triggers disease resistance in tomato plants carrying Pto[6,7,12], a serine/threonine protein kinase, and Prf[8], a Pto-interacting[13] protein. It is widely thought that interaction of AvrPto with Pto stimulates Pto kinase activity and thereby initiates resistance[5,13–15]. However, direct evidence for this hypothesis is lacking *in vivo* or *in vitro*. In contrast, one recent finding indicated that AvrPto may act as a suppressor of Pto-like kinases or Pto-like receptor kinases for suppressing MAP kinase (MAPK)-mediated basal defences[16]. An important finding contributing to understanding Pto signalling showed that Pto negatively regulates Prf-mediated resistance[15]. How AvrPto relieves this inhibition and triggers disease resistance remains elusive.

To elucidate the mechanism by which AvrPto initiates disease resistance, we determined the crystal structure of the AvrPto–Pto complex (Supplementary Fig. 1 and Supplementary Table 1). AvrPto binds Pto at a 1:1 ratio (Fig. 1a) with a binding affinity of about 0.11 μM (Supplementary Fig. 2). The interaction is primarily mediated by the contacts of one end of an AvrPto helical bundle with one Pto loop preceding β1 (Fig. 1a, b, the first interface) and the AvrPto GINP (Gly-Ile-Asn-Pro) motif with the Pto P+1 loop (Fig. 1a, c, the second interface). Compared to its solution structure[17], AvrPto in the complex remains essentially unchanged except that the GINP motif becomes well-ordered on binding to Pto (Supplementary Fig. 3). Dali search identified three kinases, cAMP-dependent protein kinase A (PKA), TGF-β receptor I (TβRI) and check point kinase 1 (CHK1) as the closest structural homologues to Pto.

Hydrophobic contacts primarily mediate the interaction around the first interface. Pto(V51) makes hydrophobic contacts with residues AvrPto(Y89, M90, L101, P102) (Fig. 1b), the importance of which was confirmed by the binding assay (Supplementary Fig. 4). In addition to Pto(H49, V51), Pto(F52) makes van der Waals contacts with AvrPto and their importance in AvrPto–Pto interaction was corroborated by mutational analysis (Fig. 1d). Fen, a closely

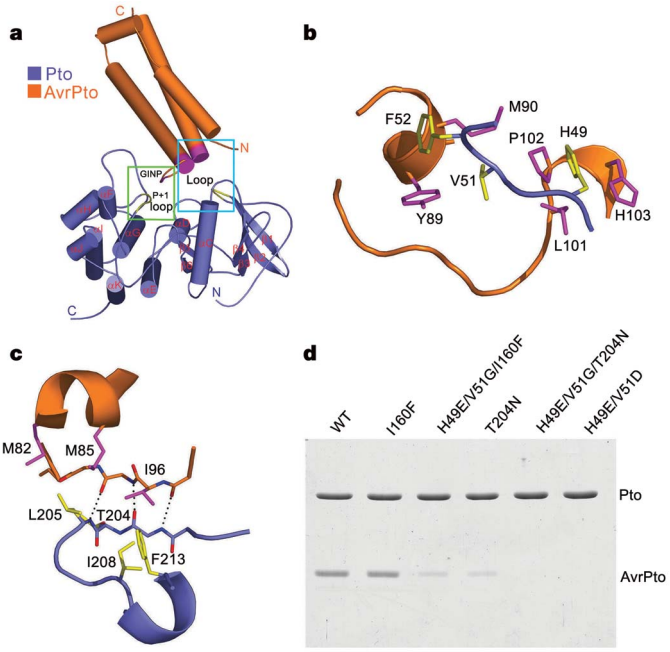


Figure 1 | Bipartite AvrPto–Pto interfaces. a, Overall structure of the AvrPto–Pto complex. Yellow, the Pto P+1 loop and the loop preceding β-1; magenta, the AvrPto GINP motif and its end of helical bundle. The first and second interfaces are highlighted in cyan and green frames, respectively. **b**, Detailed interactions between the loop preceding β-1 in Pto and the end of helical bundle of AvrPto (the first interface). **c**, Detailed interactions between the Pto P+1 loop and the AvrPto GINP motif (the second interface). **d**, Effects of point mutations in Pto on the interaction with AvrPto. WT, wild type.

[1]National Institute of Biological Sciences, No. 7 Science Park Road, Beijing 102206, China. [2]Institute of Biophysics, [3]Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China. [4]Cornell High-Energy Synchrotron Source, Cornell University, Ithaca, New York 14853, USA. [5]Department of Biological Sciences and Biotechnology, Tsinghua University, 100084, Beijing, China.
*These authors contributed equally to this work.

related Pto family member, contains Glu and Gly at H49 and V51 (Supplementary Fig. 5), respectively, and does not interact with AvrPto[18]. Incorporation of these two substitutions in Pto always resulted in a spontaneous mutation elsewhere in the protein probably owing to its toxicity to *Escherichia coli*. Nonetheless, one of these resultant mutants, Pto(H49E/V51G/I160F), was significantly compromised in its interaction with AvrPto (Fig. 1d).

Both hydrogen bonds and hydrophobic interaction contribute to the second interface. Three residues of the AvrPto GINP motif interact with the Pto P+1 loop through three main chain hydrogen bonds (Fig. 1c), whereas AvrPto(I96) packs against the hydrophobic residues Pto(I208, L205, F213) (Fig. 1c). The role of the second interface in the hypersensitive response was further confirmed by *Agrobacterium*-mediated transient assay (Supplementary Fig. 6a). Pto(T204), proposed to be the structural determinant of Pto for specific recognition of AvrPto[19], exhibited a weak interaction with AvrPto when mutated to Asn (Fig. 2d). However, substitution of all three residues in Pto—Pto(H49E/V51G/T204N)—completely abolished its interaction with AvrPto (Fig. 1d), indicating that the recognition specificity of AvrPto by Pto is conferred by these three residues. In agreement with this, AvrPto had no effect on the ability of Pto(H49E/V51G/T204N) to trigger the hypersensitive response (Supplementary Fig. 6b). Although H49 is substituted with Ala in *Lycopersicon hirsutum* Pto (Supplementary Fig. 4), it still binds AvrPto[20], because the second interface still remains intact.

The P+1 loop of Pto adopts a conformation similar to those of active PKA and CHK1 but different from that of the inactive form of TβRI (Fig. 2a), indicating that Pto in the complex assumes an active conformation. Its activation segment is stabilized through hydrogen bonds and hydrophobic interactions (Fig. 2b). Like many RD kinases[21,22], which contain an arginine (R) preceding the catalytic aspartate (D) in the catalytic loop, phosphorylation of the Pto activation loop is critical in determining its active form. Both electron density around Pto(T199) (Fig. 2c) and mass spectrometry data (not shown) indicated that this residue had been phosphorylated in the crystal. As with other RD kinases (Supplementary Fig. 7), phosphorylation of Pto(T199) to p-Pto(T199) has an important role in maintaining Pto in the active conformation by forming salt bridges with Pto(R163) and Pto(K187) (Fig. 2c). Supporting a role of p-Pto(T199) in Pto kinase activity, mutation of this residue significantly compromised Pto autophosphorylation (Fig. 2d) and phosphorylation of the substrate Pti1 (Supplementary Fig. 8b). A higher level of kinase activity of the same mutant in a previous study[23] probably resulted from usage of a preferred Pto cofactor, $MnCl_2$ (Supplementary Fig. 8a). Pto(Y207) is also important for stabilizing the P+1 loop by making hydrophobic contacts with Pto(V229, I241) and the aliphatic portion of Pto(K166), and hydrogen bonding with Pto(E233, S167) (Fig. 2b).

To investigate if the active conformation of Pto is required for AvrPto–Pto interaction, we mutated residues that stabilize the P+1 loop (Fig. 2b) and examined their kinase activity and interaction with AvrPto. One of these residues, Pto(S226), immediately underneath the P+1 loop, supports the loop in the proper conformation (Fig. 2b). Pto(S226D) had completely abolished kinase activity and interaction with AvrPto (Fig. 2d). An interaction between Pto(T199A) and AvrPto was not detected by Coomassie blue staining (Fig. 2d) but by the more sensitive silver staining, indicating a weak interaction (Supplementary Fig. 9), consistent with previous study[23]. These results suggest that the phosphorylation-stabilized P+1 is important for AvrPto–Pto interaction. Consistently, Pto(R163A) disrupting a salt bridge (Fig. 2c) abolished its interaction with AvrPto in this (Fig. 2d)

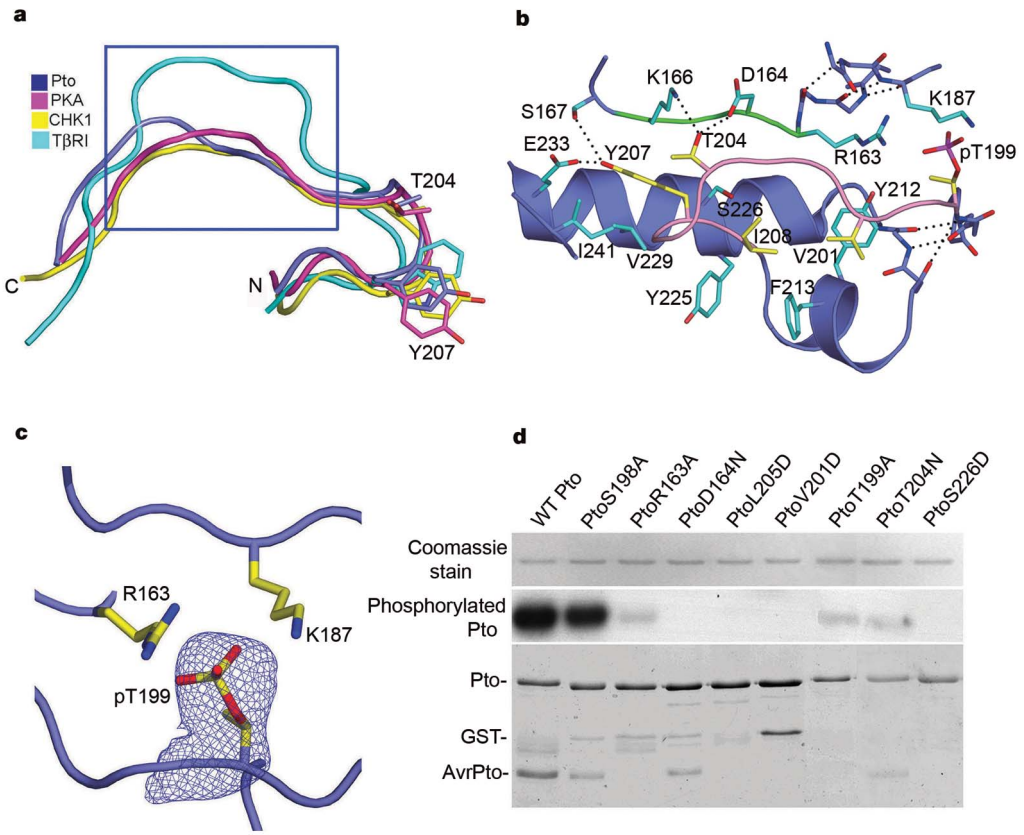


Figure 2 | The active conformation of Pto is important for AvrPto binding. **a**, Structural alignment of the Pto P+1 loop (highlighted in the blue frame) with those of kinases PKA, CHK1 and TβR1. **b**, Hydrogen bonds and hydrophobic contacts are involved in maintaining the Pto activation segment in the active conformation. The catalytic loop and activation segment of Pto are coloured in green and pink, respectively. **c**, Omit electron density map for the phosphorylated Pto(T199) (shown at 1.2 σ). **d**, Effects of various Pto mutations on kinase activity and interaction with AvrPto. The mutant Pto(L205D) (ref. 15) is a negative control. GST, glutathione *S*-transferase.

and a previous study[15]. A role of p-Pto(T199) in AvrPto-triggered hypersensitive response was supported by the observation[23] that Pto(T199A) decreased AvrPto-triggered hypersensitive response. The importance of p-Pto(T199) for AvrPto recognition is consistent with previous findings that Pto mutants abolishing the kinase activity also eliminated their interaction with AvrPto[6,7,14,15,23]. However, an exception to this has been reported. The kinase-deficient mutant Pto(D164N) still interacts with AvrPto, as shown in a previous[14] and this (Fig. 2d) study. This may have resulted from the introduction of one hydrogen bond between oxygen OD1 of Asn and the amide nitrogen of Pto(T204), thus stabilizing the Pto P+1 loop and enabling the interaction with AvrPto. Consistent with this possibility, Pto(D164A) exhibited no interaction with AvrPto[15]. To support further the important role of p-Pto(T199) in Pto recognition of AvrPto, *trans*-phosphorylation of Pto(T199) in the kinase-deficient mutant Pto(K69H) by the wild-type Pto (Supplementary Fig. 10) restored the ability of this mutant to interact with AvrPto (Supplementary Fig. 11). These results indicate that p-Pto(T199) rather than the kinase activity itself is important for Pto recognition of AvrPto.

Previous studies identified a number of constitutive gain-of-function (CGF) Pto mutants that elicit AvrPto-independent but Prf-dependent hypersensitive response[14,15]. Our structure indicates that many of these mutations destabilize the P+1 loop of Pto. In addition, all of the CGF mutants in the P+1 loop abolished or greatly reduced Pto kinase activity (Supplementary Fig. 12). These results suggest that proper conformation around the Pto P+1 loop region is important for inhibition of Prf-dependent hypersensitive response. To verify this, we tested CGF hypersensitive response activity of two Pto derivatives—Pto(S226D) and Pto(V201D)—carrying substitutions in residues that stabilize the P+1 loop from outside (Fig. 2b). These two mutants exhibited no kinase activity *in vitro* and had impaired interactions with AvrPto (Fig. 2d). As expected, both mutants exhibited strong CGF hypersensitive response activity (Fig. 3a). Contrary to Pto(Y207D) (ref. 14), the mutant Pto(Y207F), like Pto(Y207W) (ref. 14), did not exhibit CGF hypersensitive response activity (Fig. 3c), because Phe can still form hydrophobic contacts with Pto(V229, I241) and the aliphatic portion of Pto(K166) (Fig. 2b), thus stabilizing the P+1 loop.

To examine the *in vivo* importance of the first interface in elicitation of a hypersensitive response, we generated mutants Pto(H49E/V51G, H49D/V51D) and transiently expressed them in *Nicotiana benthamiana*. Strikingly, these two mutants also elicited a CGF hypersensitive response (Fig. 3b), indicating that, like the P+1 loop, this loop also exerts an inhibitory effect on the hypersensitive response. In full support of the *in vivo* importance of this interface, overexpression of Fen in *N. benthamiana* phenocopies the hypersensitive response induced by these two mutants[18]. As with CGF mutants in the P+1 loop, these two mutants were also significantly compromised in their interaction with AvrPto (Fig. 1d).

Structural comparison reveals that the interaction of Pto with the AvrPto GINP motif is similar to that of PKA with its pseudosubstrate PKI (Fig. 4a), a peptide inhibitor of PKA[24], arguing against the possibility that AvrPto promotes Pto kinase activity. To test this, we examined the kinase activity of Pto in the presence of AvrPto. Pto autophosphorylation (Fig. 4b) and phosphorylation towards a substrate Pti1 (Fig. 4c) were significantly reduced in the presence of AvrPto, with a half-maximal inhibition of Pti1 phosphorylation at about 11 μM (Supplementary Fig. 13). The inhibitory effect apparently resulted from AvrPto–Pto interaction, because AvrPto(Y89D) (Fig. 4d) and Pto(H49D/V51D) (Fig. 4e), which do not interact with their wild-type partners, showed little impact on Pto kinase activity, indicating that binding of AvrPto consequently inhibits the Pto kinase activity.

Defences are unlikely to be initiated by AvrPto inhibition of Pto kinase, because Pto(H49D/V51D), which is insensitive to kinase inhibition by AvrPto (Fig. 4e), constitutively activates the hypersensitive response in plants (Fig. 3b). It is logical to conclude that it is the AvrPto–Pto interaction, rather than the altered kinase activity, that activates Prf. Mutations that disrupt proper conformation around the P+1 loop region can relieve its inhibition and generate a CGF hypersensitive response without the requirement for Pto kinase activity (Fig. 3a and ref. 15). However, Pto kinase activity is required for AvrPto-triggered resistance[6,7,14,15,23]. In addition to recognizing AvrPto (Fig. 2d), Pto kinase activity seems to be critical downstream of recognition for induction of disease resistance, because the autophosphorylation site Pto(S198) was shown to be required for elicitation of a hypersensitive response (ref. 23) but dispensable for interaction with AvrPto (Fig. 2d and ref. 23). It is possible that p-Pto(S198) is involved in Pto conformational changes critical for Pto signalling[13] (Supplementary Fig. 14). A very recent study showed that Prf and Pto form a complex *in vivo* in which Pto probably acts as a switch for hypersensitive response induction[13]. The two AvrPto-interacting loops in Pto negatively regulate the hypersensitive response (Fig. 3), probably by direct or indirect interaction with Prf. On the other hand, Pto apparently plays a positive part in activating Prf, because plants lacking Pto do not trigger disease resistance[5]. We propose that the two loops of Pto act as a lock to keep Prf in the inactive state, whereas binding of AvrPto unlocks them and allows conformational changes in Pto, altering the way Pto interacts with Prf and consequently activating Prf (Supplementary Fig. 14).

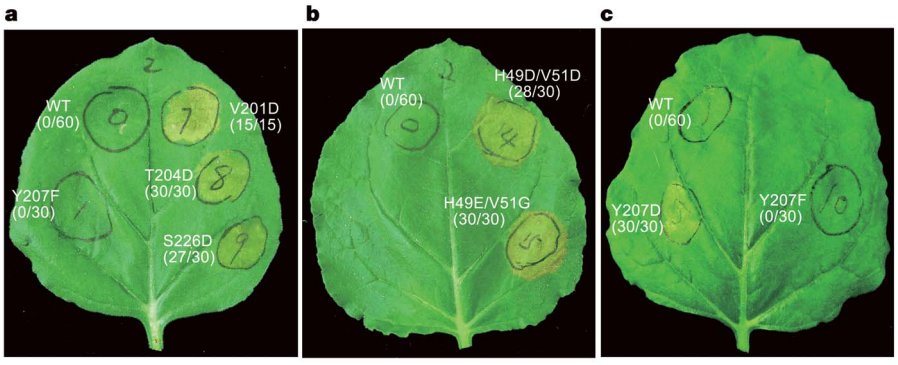


Figure 3 | Mutations in the two AvrPto–Pto interfaces trigger a CGF hypersensitive response in *N. benthamiana*. **a**, Transient expression of Pto mutants—Pto(Y207F, V201D, T204D, S226D)—with mutations in the residues that are important for maintaining the conformation of P+1 in *Nicotiana benthamiana*. The mutant Pto(T204D) is a positive control (ref. 15). **b**, Transient expression of Pto mutants Pto(H49D/V51D) and Pto(H49E/V51G) with mutations in the residues that are important for the AvrPto–Pto interaction in the first interface (Fig. 1b). **c**, Transient expression of Pto(Y207) mutants Pto(Y207F, Y207D) that stabilize the P+1 loop through hydrophobic contacts (Fig. 2b). The numbers in brackets indicate the number of leaves showing the symptoms of a hypersensitive response divided by the total number of leaves infiltrated with the indicated plasmid.

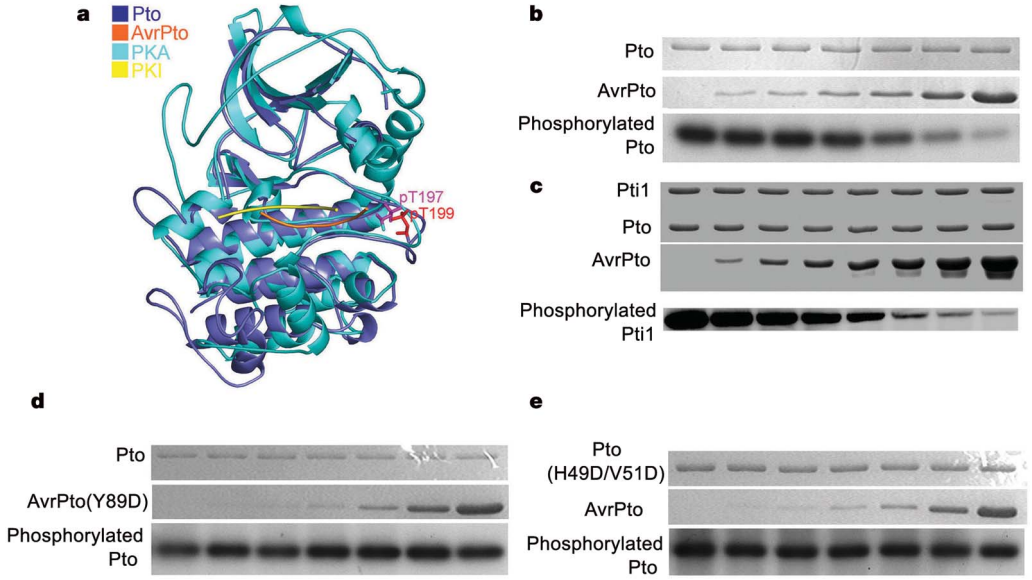


Figure 4 | AvrPto inhibits the kinase activity of Pto *in vitro*. **a**, The binding mode of the AvrPto–Pto complex is similar to that of the PKA–PKI complex. Structural alignment of the AvrPto–Pto complex with PKA–PKI complex. Only the GINP loop is shown in AvrPto. **b**, The Pto autophosphorylation activity is inhibited by AvrPto *in vitro*. **c**, Pto phosphorylation of Pti1 is inhibited by AvrPto *in vitro*. Pti1(K96Q) was used as the substrate of Pto. **d**, **e**, Mutations in AvrPto (**d**) and Pto (**e**) that disrupt the AvrPto–Pto interaction impair the kinase inhibition activity of AvrPto. Longer exposure was used for panel **e** than for panel **b**.

Inhibition of the Pto kinase activity by AvrPto is not the signal to initiate a hypersensitive response. However, the kinase inhibition activity of AvrPto can be important for its virulence function. This possibility gains support from a recent study showing that AvrPto and AvrPtoB act as suppressors of an early signalling component(s) upstream of the MAPK cascade that is activated on perception of the bacterial flagellar peptide by the receptor kinase FLS2 in *Arabidopsis*[16]. Suppression of MAPK signalling by AvrPto can be relieved by overexpression of Pto and FLS2. Therefore, Pto-like kinases or Pto-like receptor kinases were suggested to be the virulence targets of AvrPto or AvrPtoB[16]. Functional mimicry of host proteins is an important mechanism used by microbial pathogens to modulate host cellular functions[25]. Accordingly, it is conceivable that some host proteins may have evolved to deceive pathogens by mimicking their virulence targets and then initiating defences when targeted. This explains why the active conformation of Pto is required for interaction with AvrPto (Fig. 2).

METHODS SUMMARY

Protein expression, protein–protein interaction and phosphorylation assays. AvrPto (residues 29–131) in pET30 (Novagen) and the full-length Pto in pGEX-2T (Pharmacia) were co-expressed in *E. coli* strain BL21(DE3). The soluble fraction of the AvrPto–Pto complex was purified using an affinity column and further cleaned by anion-exchange column and gel filtration chromatography. Size exclusion chromatography was employed to detect AvrPto–Pto interaction. Aliquots of peak fraction corresponding to the position of AvrPto–Pto complex were subjected to SDS-polyacrylamide gel electrophoresis. The proteins were visualized by Coomassie blue staining. Buffer containing 20 mM Tris-HCl (pH 7.2), 2 mM dithiothreitol, 5 mM $MgCl_2$ and 10 μM ATP, 2 μCi [γ-^{32}P] (5,000 Ci mmole^{-1}) was used for Pto autophosphorylation and phosphorylation of the substrate Pti1. All the reactions were incubated at 30 °C for 30 min and terminated by adding an equal volume (50 μl) of 2× SDS buffer. SDS-polyacrylamide gel electrophoresis was used to fractionate proteins, and the phosphorylated proteins were visualized using a phosphoimager.

Crystallography. Crystals of the AvrPto–Pto complex were grown using hanging drop vapour diffusion. The native (3.2 Å) and MAD (4.0 Å) data sets for these crystals were collected at the BSRF (Beijing, China) beam line 3W1A using a CCD detector and processed using the software DENZO and Scalepack[26]. The AvrPto–Pto crystal structure was determined by MAD. The ordered selenium sites were positioned and refined by SOLVE/RESOLVE[27]. The experimental electron density was sufficient for model-building with the program O[28] and structure refinement with REFMAC5 (ref. 29). Statistics are given in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 5 July; accepted 24 July 2007.
Published online 12 August 2007.

1. Flor, H. H. Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275–296 (1971).
2. Belkhadir, Y., Subramaniam, R. & Dangl, J. L. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr. Opin. Plant Biol.* **7**, 391–399 (2004).
3. Chisholm, S. T., Coaker, G., Day, B. & Staskawicz, B. J. Host–microbe interactions: shaping the evolution of the plant immune response. *Cell* **124**, 803–814 (2006).
4. Innes, R. W. Guarding the goods. New insights into the central alarm system of plants. *Plant Physiol.* **135**, 695–701 (2004).
5. Pedley, K. F. & Martin, G. B. Molecular basis of Pto-mediated resistance to bacterial speck disease in tomato. *Annu. Rev. Phytopathol.* **41**, 215–243 (2003).
6. Scofield, S. R. *et al.* Molecular basis of gene-for-gene specificity in bacterial speck disease of tomato. *Science* **274**, 2063–2065 (1996).
7. Tang, X. *et al.* Initiation of plant disease resistance by physical interaction of AvrPto and Pto kinase. *Science* **274**, 2060–2063 (1996).
8. Salmeron, J. M. *et al.* Tomato Prf is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the Pto kinase gene cluster. *Cell* **86**, 123–133 (1996).
9. He, S. Y., Nomura, K. & Whittam, T. Type III secretion in mammalian and plant pathogens. *Biochim. Biophys. Acta* **1694**, 181–206 (2004).
10. Galan, J. E. *Salmonella* interactions with host cells: type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53–86 (2001).
11. Hueck, C. J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**, 379–433 (1998).
12. Kim, Y. J., Lin, N. C. & Martin, G. B. Two distinct *Pseudomonas* effector proteins interact with the Pto kinase and activate plant immunity. *Cell* **109**, 589–598 (2002).
13. Mucyn, T. S. *et al.* The tomato NBARC-LRR protein Prf interacts with Pto kinase *in vivo* to regulate specific plant immunity. *Plant Cell* **18**, 2792–2806 (2006).
14. Rathjen, J. P., Chang, J. H., Staskawicz, B. J. & Michelmore, R. W. Constitutively active Pto induces a Prf-dependent hypersensitive response in the absence of AvrPto. *EMBO J.* **18**, 3232–3240 (1999).
15. Wu, A. J., Andriotis, V. M., Durrant, M. C. & Rathjen, J. P. A patch of surface-exposed residues mediates negative regulation of immune signaling by tomato Pto kinase. *Plant Cell* **16**, 2809–2821 (2004).
16. He, P. *et al.* Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in *Arabidopsis* innate immunity. *Cell* **125**, 563–575 (2006).
17. Wulf, J., Pascuzzi, P. E., Fahmy, A., Martin, G. B. & Nicholson, L. K. The solution structure of type III effector protein AvrPto reveals conformational and dynamic features important for plant pathogenesis. *Structure* **12**, 1257–1268 (2004).

18. Chang, J. H. *et al.* Functional analyses of the Pto resistance gene family in tomato and the identification of a minor resistance determinant in a susceptible haplotype. *Mol. Plant Microbe Interact.* **15,** 281–291 (2002).
19. Frederick, R. D., Thilmony, R. L., Sessa, G. & Martin, G. B. Recognition specificity for the bacterial avirulence protein AvrPto is determined by Thr-204 in the activation loop of the tomato Pto kinase. *Mol. Cell* **2,** 241–245 (1998).
20. Riely, B. K. & Martin, G. B. Ancient origin of pathogen recognition specificity conferred by the tomato disease resistance gene *Pto*. *Proc. Natl Acad. Sci. USA* **98,** 2059–2064 (2001).
21. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109,** 275–282 (2002).
22. Nolen, B., Taylor, S. & Ghosh, G. Regulation of protein kinases; controlling activity through activation segment conformation. *Mol. Cell* **15,** 661–675 (2004).
23. Sessa, G., D'Ascenzo, M. & Martin, G. B. Thr38 and Ser198 are Pto autophosphorylation sites required for the AvrPto–Pto-mediated hypersensitive response. *EMBO J.* **19,** 2257–2269 (2000).
24. Bossemeyer, D., Engh, R. A., Kinzel, V., Ponstingl, H. & Huber, R. Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn^{2+} adenylyl imidodiphosphate and inhibitor peptide PKI(5–24). *EMBO J.* **12,** 849–859 (1993).
25. Stebbins, C. E. & Galan, J. E. Maintenance of an unfolded polypeptide by a cognate chaperone in bacterial type III secretion. *Nature* **414,** 77–81 (2001).
26. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
27. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
28. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1997).
29. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53,** 240–255 (1997).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Innes, S. He and X. Tang for critical reading and comments on our manuscript, and Y. Dong and P. Liu at the BSRF (Beijing, China) beam line 3W1A for assistance with the data collection. We are grateful to X. Liu and L. Ma for help with SPR assay. This research is funded by a Chinese Ministry of Science and Technology grant to J.C. and to J.-M.Z.

Author Contributions W.X. purified, crystallized and determined the structure and performed biochemical assays; Y.Z. performed *Agrobacterium*-mediated transient expression; Q.L., Q. Huang and Q. Hao determined structure; J.L. and X.L. purified proteins; S.C. performed the mass spectrometry assay; J.-W.W. measured the half-maximal inhibitory concentration; R.B. and L.Z. were involved in the study design; and J.-M.Z. and J.C. designed the study, analysed data and prepared the manuscript.

Author Information The atomic coordinates and structure factors of the AvrPto–Pto complex have been deposited in the RCSB Protein Data Bank under accession code 2QKW. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.C. (chaijijie@nibs.ac.cn).

METHODS

Protein preparation. Cells expressing AvrPto and Pto were induced with 1 mM IPTG for 10 h at room temperature. Cells were collected, pelleted and then resuspended in buffer A (50 mM Tris, pH 8.0, 100 mM NaCl), supplemented with protease inhibitors. The cells were lysed by sonication and then centrifuged at 14,000 r.p.m. for 1 h. The soluble fraction of the AvrPto–Pto complex from co-expression was purified using Glutathione Sepharose 4B (GS4B) and further cleaned using an anion-exchange column (Source-15Q, Pharmacia) and gel filtration chromatography after removal of GST (Superdex200, Pharmacia). A similar method was used to purify the free forms of AvrPto and Pto, including their various mutants. The unique NCBI identifiers for the proteins are: Pto, A49332; AvrPto, AAA25728; Prf, AAC49408; Pti1, AAC61805; PKA, P36887; PKI, NP_006814; CHK1, AAC51736; TβRI, AAH71181.

Crystallization and data collection. Crystallization conditions for the AvrPto–Pto complex were determined from the sparse matrix screen (Hamptonresearch). Screening was done using hanging drop vapour diffusion by combining 2 μl of protein solution. The buffer containing 0.2 M potassium sodium tartrate, 0.1 M tri-sodium citrate, pH 5.6, and 2.0 M ammonium sulphate generated crystals, which were further optimized by adding 10 mM taurine (2-aminoethanesulfonic acid). Crystals grew to their maximum size ($0.3 \times 0.4 \times 0.5\ mm^3$) approximately within two days. The crystals were transferred to the mother liquor, containing 25% glycerol, then flash cooled in liquid nitrogen. The multiple anomalous dispersion (MAD) data sets from the selenium crystal were collected to 4.0 Å. The crystal belongs to space group $P2_12_12_1$ with cell dimensions $a = 75.465$ Å, $b = 94.591$ Å, $c = 98.741$ Å, and contains one AvrPto–Pto complex per asymmetric unit.

Structure determination and refinement of the AvrPto–Pto complex. All the residues except the first 29 residues of Pto were built for the AvrPto–Pto complex. When justified by the electron density, the phosphate group in the phosphorylated T199 in Pto was included. The final atomic model of AvrPto–Pto was refined to a crystallographic R_{work} of 27.1% and an R_{free} of 30.3% to 3.2 Å. There were no outliers in the Ramachandran plot (79.1%, 16.3%, 4.6% in the core, allowed and generously allowed regions, respectively).

Gel filtration assay for protein–protein interaction. The Pto and AvrPto proteins purified by affinity chromatography and anion-exchange column (Source-15Q, Pharmacia) were used for interaction assay. To examine the interaction between Pto and AvrPto, size exclusion chromatography (Superdex-200 column, Pharmacia Biotech) was employed. In all test runs, AvrPto and Pto were mixed together and incubated at 4 °C for 1 h. Buffer containing 25.0 mM Tris, pH 8.0, 100 mM NaCl, 3 mM dithiothreitol was used for gel filtration assays, with the flow rate of 0.5 ml min^{-1}, unless indicated otherwise. Aliquots of peak fraction corresponding to the position of the AvrPto–Pto complex were subjected to SDS-polyacrylamide gel electrophoresis. The proteins were visualized by Coomassie blue staining.

***In vitro* kinase assay.** For autophosphorylation, 1.5 μg of Pto was diluted to 50 μl using the reaction buffer. For phosphorylation of Pti1(K96Q), 1.5 μg of Pto and 5.0 μg of Pti1(K96Q) were mixed in 50 μl of reaction buffer. To determine the effect of AvrPto on Pto kinase activity, the autophosphorylation and Pti1 phosphorylation assays were carried out in the presence of different concentration of AvrPto (0.8 μg, 1.0 μg, 1.2 μg, 2 μg, 4 μg, 8 μg).

Plasmid constructs and *Agrobacterium*-mediated transient expression. For Pto expression, the *Pto* promoter was PCR-amplified from PtoR tomato plants, digested with *Sac*I and *Eco*RI, and inserted into pFAST (a gift from Y. Xia). The wild-type and mutant *Pto* sequences were PCR amplified, digested with *Bam*HI and *Xho*I, and inserted into the plasmid containing the *Pto* promoter. For AvrPto transient expression, the wild-type and mutant *AvrPto* sequences were PCR-amplified, digested with *Nde*I and *Xho*I, and inserted into the pBTEX plasmid. *Agrobacterium tumefaciens* strain GV3101 containing the binary plasmid of interest was grown in LB media with appropriate antibiotics to stationary phase at 28 degrees. A tenfold dilution was made for the cell cultures with fresh LB plus antibiotics, 10 mM MES, pH 5.6, and 50 μM acetosyringone. These diluted cells were allowed to grow overnight at 28 °C, pelleted and washed once using infiltration medium (MS medium containing 10 mM MES, pH 5.6, and 150 μM acetosyringone). Finally, the cells were resuspended to an OD_{600} of 0.5 using the same medium and injected into *Nicotiana benthamiana* six- to seven-week-old plants. To allow the development of a hypersensitive response, these plants were covered and left in the growth chamber for 4 days.

LETTERS

Structure of Dnmt3a bound to Dnmt3L suggests a model for *de novo* DNA methylation

Da Jia[1]*, Renata Z. Jurkowska[2]*, Xing Zhang[1], Albert Jeltsch[2] & Xiaodong Cheng[1]

Genetic imprinting, found in flowering plants and placental mammals, uses DNA methylation to yield gene expression that is dependent on the parent of origin[1]. DNA methyltransferase 3a (Dnmt3a) and its regulatory factor, DNA methyltransferase 3-like protein (Dnmt3L), are both required for the *de novo* DNA methylation of imprinted genes in mammalian germ cells. Dnmt3L interacts specifically with unmethylated lysine 4 of histone H3 through its amino-terminal PHD (plant homeodomain)-like domain[2]. Here we show, with the use of crystallography, that the carboxy-terminal domain of human Dnmt3L interacts with the catalytic domain of Dnmt3a, demonstrating that Dnmt3L has dual functions of binding the unmethylated histone tail and activating DNA methyltransferase. The complexed C-terminal domains of Dnmt3a and Dnmt3L showed further dimerization through Dnmt3a–Dnmt3a interaction, forming a tetrameric complex with two active sites. Substitution of key non-catalytic residues at the Dnmt3a–Dnmt3L interface or the Dnmt3a–Dnmt3a interface eliminated enzymatic activity. Molecular modelling of a DNA–Dnmt3a dimer indicated that the two active sites are separated by about one DNA helical turn. The C-terminal domain of Dnmt3a oligomerizes on DNA to form a nucleoprotein filament. A periodicity in the activity of Dnmt3a on long DNA revealed a correlation of methylated CpG sites at distances of eight to ten base pairs, indicating that oligomerization leads Dnmt3a to methylate DNA in a periodic pattern. A similar periodicity is observed for the frequency of CpG sites in the differentially methylated regions of 12 maternally imprinted mouse genes. These results suggest a basis for the recognition and methylation of differentially methylated regions in imprinted genes, involving the detection of both nucleosome modification and CpG spacing.

In both flowering plants and placental mammals, DNA methylation has a central role in imprinting, but in neither case is it clear how imprinted genes are targeted for methylation. Imprinted genes in mammals are often associated with differentially methylated regions (DMRs)[3], which show DNA methylation patterns that depend on the parent of origin. How the imprinting machinery recognizes DMRs is unknown. The Dnmt3 family includes three members: two *de novo* CpG methyltransferases, namely Dnmt3a and Dnmt3b (ref. 4), and an enzymatically inactive paralogue, Dnmt3L, that functions as a regulatory factor in germ cells[5]. Inactivating both Dnmt3a and Dnmt3b abolishes *de novo* methylation in mouse embryos[4]. Although Dnmt3b conditional germline knockout animals and their offspring show no apparent phenotype, the phenotype of a corresponding Dnmt3a conditional knockout[6] is indistinguishable from that of Dnmt3L knockout mice[5,7] with altered sex-specific *de novo* methylation of DNA sequences in male and female germ cells. These results indicate that Dnmt3a and Dnmt3L are both required for the methylation of most imprinted loci in germ cells.

We undertook structural and biochemical studies of a homogeneous complex of Dnmt3L and Dnmt3a2, generated by a co-expression and co-purification system (Supplementary Fig. 1). Dnmt3a2 is a shorter isoform of Dnmt3a that is the predominant form in embryonic stem cells and embryonal carcinoma cells and can also be detected in testis, ovary, thymus and spleen[8]. For crystallography, we focused on a stable complex of the C-terminal domains from both proteins (Dnmt3a-C and Dnmt3L-C) that retains substantial methyltransferase activity (Supplementary Fig. 1)[9]. The structure of the C-terminal complex was determined to a resolution of 2.9 Å in the presence of cofactor product *S*-adenosyl-L-homocysteine (AdoHcy) (Supplementary Table 1).

Both Dnmt3a-C and Dnmt3L-C have the classical fold characteristic for *S*-adenosyl-L-methionine (AdoMet)-dependent methyltransferases[10], but AdoHcy was found only in Dnmt3a-C, not in Dnmt3L-C (Fig. 1a). This is consistent with Dnmt3a-C being the catalytic component of the complex, whereas Dnmt3L is inactive and unable to bind cofactor[9]. The overall complex is elongated (about $160 \times 60 \times 50$ Å^3) with a butterfly shape (Fig. 1a, b). The complex contains two monomers of Dnmt3a-C and two of Dnmt3L-C, forming a tetramer (Dnmt3L–Dnmt3a–Dnmt3a–Dnmt3L) with two Dnmt3L–Dnmt3a interfaces (about 906 Å^2 interface area) and one Dnmt3a–Dnmt3a interface (about 944 Å^2). The Dnmt3L–Dnmt3a interface of Dnmt3L also supports a Dnmt3L homodimer (Supplementary Fig. 2a, b). Dnmt3a2 might use the same interface to form a Dnmt3a2 homo-oligomer, as suggested by analytical size exclusion chromatography (a broad peak of about 500 kDa; Supplementary Fig. 2c, d). An F728A mutation of Dnmt3a2, which eliminates a hydrophobic interaction at the Dnmt3a–Dnmt3L interface, disrupted the Dnmt3a2 homo-oligomer to yield a roughly 150-kDa dimer (Supplementary Fig. 2c, d; the calculated mass of a Dnmt3a2 monomer is 78 kDa) and abolished methyltransferase activity (Fig. 1c; compare lanes 1 and 4). The equivalent mutant in Dnmt3L, F261A, lost its ability to form a homodimer (Supplementary Fig. 2a, b) and simultaneously its ability to stimulate wild-type Dnmt3a2 activity (Fig. 1c; compare lanes 1–3). At the Dnmt3a–Dnmt3a interface, an R881A mutation of Dnmt3a that eliminates a network of polar interactions (Fig. 1d) abolished the activity of Dnmt3a-C (ref. 11). These data indicate that both interfaces (Dnmt3a–Dnmt3L and Dnmt3a–Dnmt3a) are essential for catalysis. Dnmt3L might stabilize the conformation of the active-site loop of Dnmt3a (residues 704–725 before helix αD, containing the key nucleophile Cys 706), by means of interactions with the C-terminal portion of the active-site loop (Supplementary Fig. 3). These stabilizing interactions could explain the stimulation of Dnmt3a2 activity by Dnmt3L (Fig. 1c; lanes 1 and 2)[9,12–14], as well as the linked loss of the Dnmt3a–Dnmt3L interface and of catalytic activity in Dnmt3a2 F728A (Fig. 1c, lanes 4–6).

[1]Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322, USA. [2]Biochemistry Laboratory, School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany.
*These authors contributed equally to this work.

Among the known active DNA methyltransferases, Dnmt3a and Dnmt3b have the smallest DNA-binding domain (absent from Dnmt3L; Supplementary Fig. 4). This domain includes about 50 residues in Dnmt3a/Dnmt3b in comparison with, for example, about 85 residues in the bacterial GCGC methyltransferase M.*Hha*I (ref. 15). However, dimerization by means of the Dnmt3a–Dnmt3a interface brings two active sites together and effectively doubles the DNA-binding surface. We superimposed the Dnmt3a structure on that of M.*Hha*I complexed with a short oligonucleotide[15]. This yielded a model of a Dnmt3a–DNA complex with a short DNA duplex bound to each active site (Fig. 2a). The two DNA segments can be connected easily to form a contiguous DNA, such that the two active sites are located in the major groove about 40 Å apart (Fig. 2b). This model indicates that dimeric Dnmt3a could methylate two CpGs separated by one helical turn in one binding event.

Electrophoretic mobility-shift assays revealed cooperative multimerization on DNA of Dnmt3a-C alone or of the Dnmt3a-C–Dnmt3L-C complex, with each monomer of Dnmt3a-C binding to about 12 base pairs (Fig. 2c and Supplementary Fig. 5a). Gel-filtration experiments, using short oligonucleotides of different lengths, confirmed the oligomerization of Dnmt3a-C on DNA with one monomer bound for each roughly nine base pairs (Supplementary Fig. 5b). Oligonucleotides containing a single CpG site are substrates for the Dnmt3a–Dnmt3L C-terminal complex, but at least eight base pairs on each side of the CpG are required for substantial activity, which is consistent with a possible requirement for DNA contact by both Dnmt3a molecules (Supplementary Fig. 1d).

On longer DNAs, we tested the possibility that the Dnmt3a dimer, in one binding event, methylates two CpG sites separated by a helical turn. Two different DNA fragments were methylated *in vitro* by Dnmt3a-C. Methylation was analysed by bisulphite conversion, followed by cloning and sequencing, of 119 clones in total. At an overall methylation level of 22–26% there was a periodic fluctuation of the relative methylation at the various CpG sites (Fig. 3a and Supplementary Fig. 6). To determine whether there is a correlation between the methylation states of any two CpG sites at a given distance from one another, the autocorrelation of the methylation states was calculated for all pairs of CpGs in each individual clone. We observed a highly significant correlation of methylation status at distances of eight to ten base pairs between two CpG sites (Fig. 3b). These experiments were performed under conditions in which the DNA was saturated with the enzyme. As a result of its large interface with DNA, the enzyme oligomer or polymer cannot move along the DNA, in agreement to the observation that Dnmt3a-C methylates DNA in a non-processive manner[16]. The enzyme oligomer on the DNA presents the active sites in a regular spacing, which leads to a correlated methylation of CpG sites. In contrast, CpG sites positioned between the active sites are not readily available for methylation, which causes a correlation of absence of methylation that has the same period (Supplementary Fig. 6c).

The similarity of defects observed in the Dnmt3a conditional germline knockout and the Dnmt3L-null mutants indicates that both Dnmt3a and Dnmt3L are required for the methylation of DMRs in imprinted genes[5–7]. We studied the distribution of CpG sites among 12 known maternally imprinted DMRs[17] that are methylated in wild-type embryos and responsible for their germline targeting, including the three (*Snrpn*, *Igf2r* and *Peg1*) that were shown experimentally to be unmethylated in affected embryos[6]. The frequencies of the distances between CpG sites peak periodically, with an average interval of 9.5 base pairs (Fig. 4a; the λ DNA fragment and the CpG island used as methylation substrates in Fig. 3 do not contain such pattern). The periodic occurrence of CpG sites 9.5 base pairs apart on average (examples are shown in Fig. 4d) makes these DNA sequences an ideal

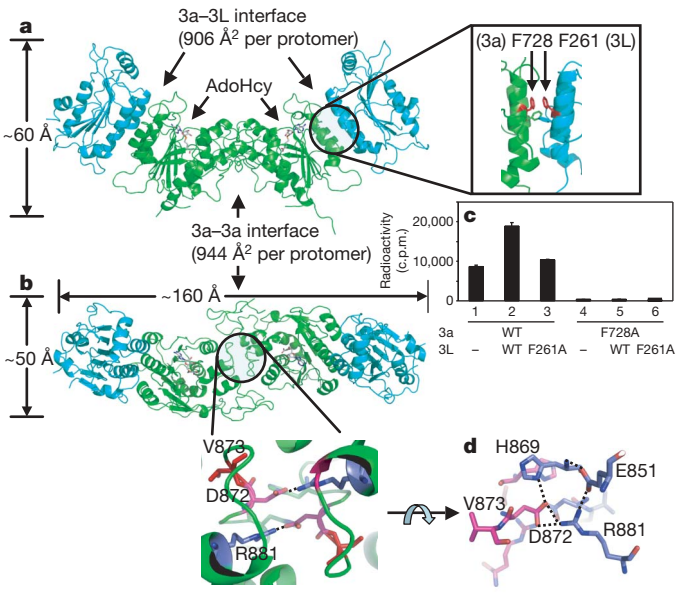


Figure 1 | Structure of the Dnmt3a–Dnmt3L (3a–3L) C-terminal domain complex. **a**, Side view of the tetramer Dnmt3L (blue)–Dnmt3a (green)–Dnmt3a (green)–Dnmt3L (blue), with the bound AdoHcy shown as a stick model. The inset shows the two pairs of phenylalanine residues (F728 and F768 of Dnmt3a, and F261 and F301 of Dnmt3L) in the centre of the Dnmt3a–Dnmt3L interface. **b**, Top view showing the Dnmt3a–Dnmt3a interface with two pairs of salt bridges formed between R881 and D872 (enlarged). **c**, Activities of Dnmt3a2 and its point mutant F728A in the presence and absence of Dnmt3L. Error bars represent s.d. calculated from two independent experiments. **d**, A network of polar interactions between two Dnmt3a molecules (coloured blue and purple) involves R881–D872, D872–H869, H869–E851 and E851–R881.

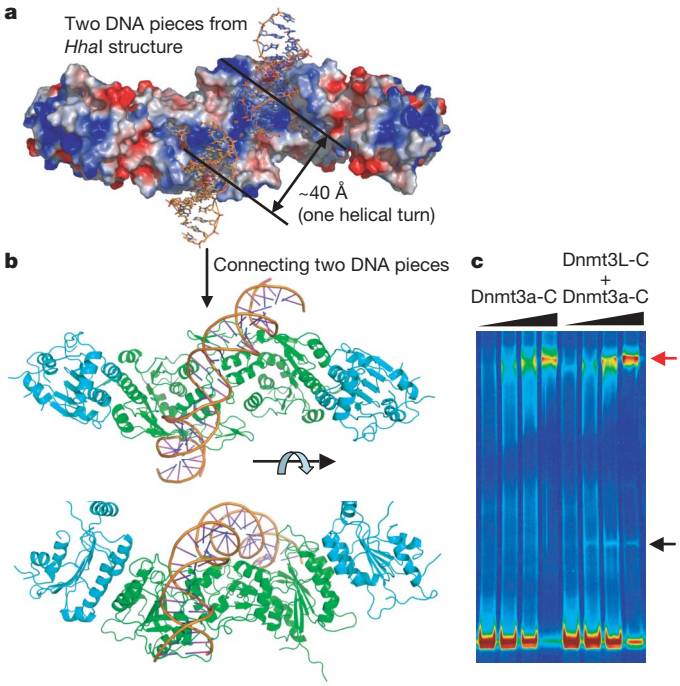


Figure 2 | Model of Dnmt3a–Dnmt3L tetramer with DNA. **a**, Surface representation of the Dnmt3a–Dnmt3L tetramer, with two short DNA molecules adopted by superimposition of the *Hha*I–DNA complex structure[15] onto individual Dnmt3a-C. **b**, Two views of the Dnmt3a–Dnmt3L tetramer with one contiguous curved DNA molecule, approximately 25 base pairs in length, covering two active sites. **c**, Cooperative formation of large protein–DNA complexes (red arrow) after incubation of a 146-base-pair DNA with increasing concentrations of Dnmt3a-C or Dnmt3a-C–Dnmt3L-C complex. The lower band (black arrow) corresponds to one Dnmt3L-C bound to the DNA as a result of the presence of some free Dnmt3L-C in the mixture (Supplementary Fig. 5a).

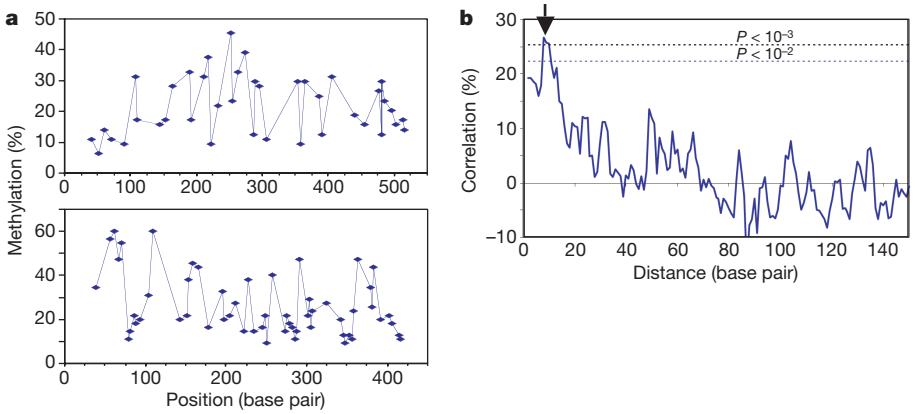


Figure 3 | Periodicity of Dnmt3a-C activity on long DNA substrates. a, Two DNA fragments (from λ phage (top; 520 base pairs, 40 CpGs) and the CpG island upstream of the human *SUHW1* gene (bottom; 420 base pairs, 56 CpGs)) were used for methylation. A total of 119 clones (64 from λ phage, 55 from the human CpG island) were sequenced to determine their methylation pattern. **b**, Autocorrelation of CpG methylation. By averaging two substrates, the *P* values for the correlation factors observed for distances of eight to ten base pairs (arrow) are in the range (3–7) $\times 10^{-4}$. This periodicity is detectable for correlation of methylation of individual CpG sites as well as for correlation of the absence of methylation at two sites (Supplementary Fig. 6).

substrate for the activity of two active sites in Dnmt3a, indicating that the CpGs on maternally imprinted DMRs could be methylated simultaneously by a Dnmt3a–Dnmt3L tetramer or oligomer (as shown in Fig. 2). The periodicity of the distribution of pairwise distances of CpGs was further analysed by calculating an autocorrelation function of the frequencies, which underscored the significance of the observation (Fig. 4b). As controls, ten CpG islands randomly taken from promoter regions of genes on human chromosome 21 (Fig. 4c and Supplementary Fig. 7) did not show any correlation in the positioning of CG sites.

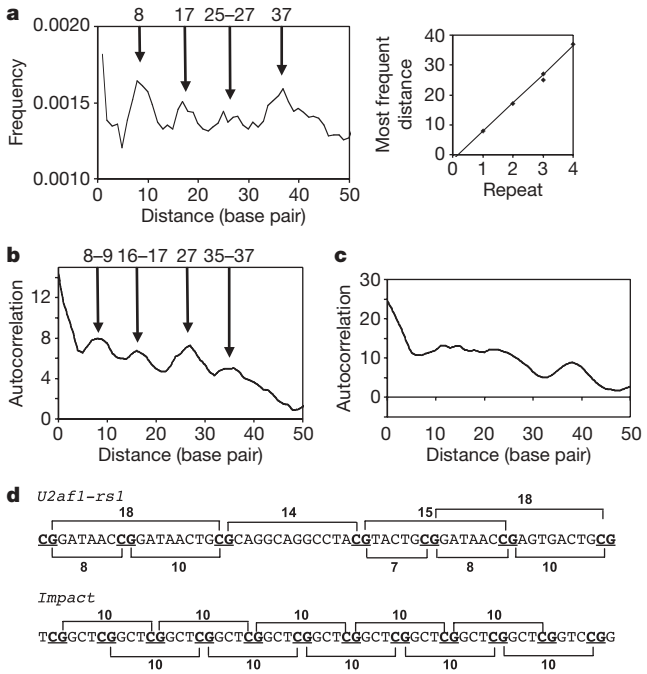


Figure 4 | Periodicity of CpG sites in maternally imprinted DMRs[17]. **a**, Frequencies of CpG pairs at a given distance with respect to all pairs, for 11 maternally imprinted DMRs. The slope of the graph in the inset is 9.5. **b**, The autocorrelation of peaks for 11 maternally imprinted DMRs. One sequence (*Impact*) was excluded from the plot because it contains repeats (see **d**), with each CpG site separated from the second subsequent one by ten base pairs; thus, it follows the same pattern seen by the other 11 DMRs, bringing the total number to 12 DMRs. **c**, Autocorrelation of peaks for ten CpG islands randomly selected from human chromosome 21. **d**, A region of DMR sequences of the *U2af1-rs1* and *Impact* genes[17].

In contrast with the maternally imprinted DMRs, three paternally imprinted DMRs (*H19*, *Dlk1–Gtl2* and *Rasgrf1*; Supplementary Fig. 8a) did not show such periodicity; only *Rasgrf1* showed a weak periodic pattern similar to maternally imprinted DMRs. The three paternal DMRs showed different methylation levels in impaired spermatogenesis: first, the DMR of *Rasgrf1* is normally methylated in both the Dnmt3a conditional mutant and Dnmt3L$^{-/-}$ as well as in wild-type males[6], but unmethylated in Dnmt3L knockout animals in a different study[18]; second, the *H19* DMR is unmethylated in both mutants[6] but showed mosaic methylation in two different studies[7,18]; and third, the DMR at *Dlk1–Gtl2* was methylated in Dnmt3L$^{-/-}$ animals but not in a Dnmt3a conditional mutant. It is possible that additional factors (such as RNA[19]) are involved in establishing paternal imprints at specific loci, including paternally imprinted retrotransposons (LINE-1 and IAP; Supplementary Fig. 8b). Dnmt3L and Dnmt3a are at present the only factors known to be required for establishing maternal imprints in germ cells. We conclude that the periodic arrangement of CpGs in maternally imprinted DMRs constitutes an environment that is favourable for methylation by the Dnmt3a–Dnmt3L tetramer, which is consistent with the tetramers having two active sites with similar spacing and might contribute to their preferential methylation in the female germ line. Comprehensive genome-wide studies will be required in the future to determine whether all CpG islands with a periodicity of eight to ten base pairs are maternally imprinted.

Finally, histone methylation has a function in epigenetic signalling in addition to DNA methylation. There have been reports of an inverse relationship between methylation of histone H3 lysine 4 (H3K4) and allele-specific DNA methylation at DMRs; that is, a lack of H3K4 methylation at the methylated allele and the presence of H3K4 methylation at the unmethylated allele[20–23]. The Dnmt3L–Dnmt3a complex structure presented here indicates a novel mechanism by which an absence of H3K4 methylation is recognized by the PHD (plant homeodomain)-like domain of Dnmt3L (ref. 2), whereas its C-terminal methyltransferase-like domain brings in the active DNA methyltransferase Dnmt3a to establish a heritable DNA methylation pattern. Hence, H3K4 methylation could protect unmethylated DMRs from DNA methylation by the Dnmt3a–Dnmt3L complex.

METHODS SUMMARY

Co-expression and co-purification of Dnmt3 proteins. Co-expression of full-length Dnm3a2 (residues 220–908 of Dnmt3a; National Center for Biotechnology Information (NCBI) accession number o88508) and Dnmt3L (NCBI accession number AAH83147) was achieved by engineering two expression

cassettes in one plasmid. Co-expression of Dnmt3a-C (residues 623–908) and Dnmt3L-C (residues 160–386) was achieved by the sequential transformation of two plasmids (pXC528 and pXC510) into *Escherichia coli* strain BL21 (DE3). Dnmt3a2 or Dnmt3a-C contained an N-terminal His_6 tag, and Dnmt3L or Dnmt3L-C was a glutathione *S*-transferase (GST) fusion protein. The protein complex was purified with the use of three-column chromatography (GSTrap HP column, Ni^{2+}-chelating column and Superdex 75; Amersham-Pharmacia). The GST tag was cleaved by thrombin.

Crystallography. By combining the Se-anomalous diffraction data (Supplementary Table 1) and the molecular replacement with the use of the C-terminal domain of Dnmt3L homodimer as the initial search model, the structure of Dnmt3a-C–Dnmt3L-C complex was solved.

Sequence analyses. The analyses of the periodicity of CpG positioning and in Dnmt3a activity were performed with two in-house programs.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 30 May; accepted 6 August 2007.
Published online 22 August 2007.

1. Feil, R. & Berger, F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet.* **23,** 192–199 (2007).
2. Ooi, S. K. T. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
3. Ferguson-Smith, A. C. & Surani, M. A. Imprinting and the epigenetic asymmetry between parental genomes. *Science* **293,** 1086–1089 (2001).
4. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99,** 247–257 (1999).
5. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536–2539 (2001).
6. Kaneda, M. *et al.* Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429,** 900–903 (2004).
7. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).
8. Chen, T., Ueda, Y., Xie, S. & Li, E. A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active *de novo* methylation. *J. Biol. Chem.* **277,** 38746–38754 (2002).
9. Gowher, H., Liebert, K., Hermann, A., Xu, G. & Jeltsch, A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J. Biol. Chem.* **280,** 13341–13348 (2005).
10. Schubert, H. L., Blumenthal, R. M. & Cheng, X. Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.* **28,** 329–335 (2003).
11. Gowher, H. *et al.* Mutational analysis of the catalytic domain of the murine Dnmt3a DNA-(cytosine C5)-methyltransferase. *J. Mol. Biol.* **357,** 928–941 (2006).
12. Chedin, F., Lieber, M. R. & Hsieh, C. L. The DNA methyltransferase-like protein DNMT3L stimulates *de novo* methylation by Dnmt3a. *Proc. Natl Acad. Sci. USA* **99,** 16916–16921 (2002).
13. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H. & Tajima, S. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J. Biol. Chem.* **279,** 27816–27823 (2004).
14. Kareta, M. S., Botello, Z. M., Ennis, J. J., Chou, C. & Chedin, F. Reconstitution and mechanism of the stimulation of *de novo* methylation by human DNMT3L. *J. Biol. Chem.* **281,** 25893–25902 (2006).
15. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76,** 357–369 (1994).
16. Gowher, H. & Jeltsch, A. Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases. *J. Biol. Chem.* **277,** 20409–20414 (2002).
17. Kobayashi, H. *et al.* Bisulfite sequencing and dinucleotide content analysis of 15 imprinted mouse differentially methylated regions (DMRs): paternally methylated DMRs contain less CpGs than maternally methylated DMRs. *Cytogenet. Genome Res.* **113,** 130–137 (2006).
18. Webster, K. E. *et al.* Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl Acad. Sci. USA* **102,** 4068–4073 (2005).
19. Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G. J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316,** 744–747 (2007).
20. Delaval, K. *et al.* Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* **26,** 720–729 (2007).
21. Fournier, C. *et al.* Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J.* **21,** 6560–6570 (2002).
22. Yamasaki, Y. *et al.* Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. *Hum. Mol. Genet.* **14,** 2511–2520 (2005).
23. Vu, T. H., Li, T. & Hoffman, A. R. Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse. *Hum. Mol. Genet.* **13,** 2233–2245 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. R. Horton for assistance with X-ray diffraction data collection; Z. Yang for assistance in solving the structure; H. Sasaki and R. Hirasawa for providing DMR sequences; S. Devine and R. Mills for help with DMR sequence analysis; A. Pingoud for providing an R.*Eco*RV expression clone used for calibration of the EMSA experiments; R. M. Blumenthal for critical editing of the manuscript; and E. Bernstein and R. E. Collins for comments on the manuscript. This work was supported by grants from the National Institutes of Health to X.C. and grants from the Deutsche Forschungsgemeinschaft and BMBF (Biofuture programme) to A.J.

Author Information The X-ray structure of Dnmt3a–Dnmt3L C-terminal tetramer complex is deposited in the Protein Data Bank under ID code 2QRV. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.J. (a.jeltsch@jacobs-university.de) or X.C. (xcheng@emory.edu).

METHODS

Expression and purification. The sequence encoding Dnmt3a2 (residues 220–908 of Dnmt3a; NCBI accession number o88508) was subcloned into the pET-28a vector (Novagen) using *Nde*I/*Eco*RI sites, yielding pXC465. *E. coli* BL21 (DE3) Codon-plus RIL (Stratagene) carrying pXC465 was grown in Luria–Bertani medium supplemented with 50 μM $ZnSO_4$ at 37 °C to $D_{600} \approx 0.6$, then shifted to 22 °C for 20 min and induced overnight with 0.4 mM isopropyl β-D-thiogalactoside. The complementary DNA encoding full-length Dnmt3L was subcloned into a modified pET-28a vector, yielding pXC391, with an N-terminal His_6 tag. Both proteins were purified separately by three-column chromatography and used in the methyl transfer assays (Fig. 1c).

Co-expression of Dnm3a2 and Dnmt3L (NCBI accession number AAH83147) was achieved by engineering two expression cassettes in one plasmid. The sequence encoding Dnmt3a2 together with a T7 promoter and the terminator was amplified, digested with *Mfe*I, and cloned into the *Eco*RI site of pGEX-2T vector (Amersham-Pharmacia) containing Dnmt3L fused to GST followed by a thrombin cleavage site. Dnmt3a2 contains an N-terminal His_6 tag, and Dnmt3L is a GST fusion. The complex was purified by a GST-affinity column, GST fusion cleavage by thrombin, and a Ni^{2+}-chelating column to select for the complex. Finally a gel-filtration column yielded a stable complex at 1:1 stoichiometry.

The sequence encoding Dnmt3a residues 623–908 (Dnmt3a-C) was amplified and cloned into a modified pACYCduet-1 vector (with an N-terminal His_6 tag) using *Nde*I/*Xho*I sites, yielding pXC528. The C-terminal residues 160–386 of Dnmt3L (Dnmt3L-C) were amplified, digested with *Bcl*I and *Eco*RI, and cloned into the *Bam*HI and *Eco*RI sites of pGEX-2T, yielding pXC510. Plasmids pXC528 and pXC510 were transformed sequentially into *E. coli* strain BL21 (DE3). The proteins were produced with the auto-induction method[24]. *E. coli* strain BL21 (DE3) carrying pXC528 and pXC510 was grown in ZY-5052 medium at 37 °C for about 16 h before harvest. The D_{600} value of harvested cells was about 6. The protein complex was purified by three-column chromatography. The GST tag was cleaved by thrombin. The complex was stored in 20 mM Tris-HCl pH 8.0, 250 mM NaCl, 5% glycerol and 0.1% 2-mercaptoethanol.

Methyl transfer assays using short oligonucleotides. The methyl transfer activity assays were performed in 25 mM Tris-HCl pH 7.5, 5% glycerol and 0.1% 2-mercaptoethanol). The reaction mixture (20 μl total volume) contained 3.7 μM [*methyl*-^{3}H]AdoMet (14.9 Ci/mmol; NEN), and 0.75 μM $(GAC)_{12}$ oligonucleotides (Fig. 1c and Supplementary Fig. 1c) or 2.25 μM single CpG-containing oligonucleotides (Fig. 2c). The mixture contained 0.3 μM (Fig. 1c and Supplementary Fig. 1c) or 0.5 μM (Fig. 2c) purified proteins. Proteins were preincubated for 5 min with AdoMet at 37 °C before the addition of DNA. The reaction were incubated at 37 °C for 1 h, and terminated by the addition of 1.5 mM unlabelled AdoMet. The reaction mixtures were spotted on DE81 paper circles (Whatman), washed five times with 2 ml of cold 0.2 M NH_4HCO_3, five times with 2 ml of deionized water and once with 2 ml of ethanol. The dried circles were subjected to liquid-scintillation counting with Cytoscint scintillant. Each reaction was performed at least in duplicate.

Dnmt3a-C activity on long DNA substrates. Two DNA fragments were used, one from λ phage comprising 520 base pairs and 40 CpGs, the other comprising 420 base pairs and 56 CpGs from the CpG island upstream of the human *SUHW1* gene. Purified polymerase chain reaction product (50–100 nM) was incubated for 3 h with 2.5 μM Dnmt3a-C (purified as described[9]). The methylation reaction was stopped by freezing in liquid N_2 and digestion with proteinase K (New England Biolabs). Afterwards the DNA was treated with sodium bisulphite essentially as described[25,26], and cloned into Topo-TA vector (Invitrogen). A total of 119 individual clones were sequenced to determine their methylation pattern.

Electrophoretic mobility-shift assays. Electrophoretic mobility-shift assay experiments were performed with a 146-base-pair polymerase chain reaction product amplified from mammary tumour virus 3′ long terminal repeat nucleosome A site[27] that was fluorescently labelled at both termini. DNA (30 nM) was incubated with protein (Dnmt3a-C alone or Dnmt3a-C and Dnmt3L-C mixture, purified as described[9,11]) in reaction buffer (20 mM HEPES pH 7.5, 1 mM EDTA, 100 mM KCl, 0.2 mM sinefungin, 0.5 mg ml^{-1} BSA) for 30 min and applied to an 8% polyacrylamide gel run in 0.5 × Tris-borate-EDTA buffer.

After incubating a 146-base-pair DNA with increasing concentrations (1, 2.5, 5 and 10 μM) of Dnmt3a-C or Dnmt3a-C–Dnmt3L-C complex, large complexes were observed (red arrow in Fig. 2c). The high degree of cooperativity is illustrated by the formation of large protein–DNA complexes and little appearance of intermediates. At the highest protein concentration used, the estimated protein mass of Dnmt3a-C bound to the DNA was 433 ± 8.5 kDa, which corresponds to about 12 protein molecules (a Dnmt3a-C monomer is 36 kDa). One Dnmt3a-C monomer therefore occupies about 12 base pairs, or a dimer occupies about 24 base pairs. This mass estimation was based on calibration with R.*Eco*RV, as detailed in Supplementary Fig. 5a.

Crystallography. The purified Dnmt3a-C–Dnmt3L-C complex was incubated with AdoHcy at a 1:4 molar ratio before concentration to 24 mg ml^{-1} (about 400 μM). The complex was incubated with short self-annealing 12–16-base-pair oligonucleotides with a single central CpG (in a molar ratio ranging from 1:1 to 1:4 excess of DNA) for at least 1 h on ice before crystallization. The final solution contained about 100–133 μM complex in 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5% glycerol and 0.1% 2-mercaptoethanol. Crystals were obtained by the hanging drop method; the mother liquor contained 2–5% PEG3000, 100 mM Tris-HCl pH 8.0 and 5% glycerol at 16 °C. It usually took a few weeks to two months before crystals reached a maximum size of about $0.05 \times 0.05 \times 0.4$ mm^3. The crystals were cryoprotected by mother liquor supplemented with 25% ethylene glycol before being plunged into liquid nitrogen. X-ray diffraction data were collected at the SER-CAT beamline at the Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois.

Using the C-terminal domain of the Dnmt3L homodimer (Supplementary Fig. 2b, top panel) as the initial search model, the molecular replacement program PHASER[28] found three solutions with a *Z* score of 17.2 in the resolution range 40–2.9 Å of native data (crystal 2 in Supplementary Table 1). Each solution represented one possible Dnmt3a–Dnmt3L heterodimer. The molecular replacement phases were then used to locate the selenium sites by anomalous difference Fourier with the selenium absorption peak data collected from crystal 1 (Supplementary Table 1). The top anomalous peaks were refined by SOLVE and RESOLVE[29], which detected a fourth heterodimer. Two pairs of Dnmt3a–Dnmt3L heterodimers were related by a two-fold symmetry that generated a tetramer. There were two tetrameric complexes per crystallographic asymmetric unit. The non-crystallographic symmetry was used as restraints in the CNS refinement[30], and was released for the side chains at the later cycles to account for different interaction environments of crystal packing with each molecule. The group *B*-factor for each residue was refined at the earlier stage of refinement, and the individual *B*-factor for each atom was refined at the later cycles. The Dnmt3L molecules have a higher *B*-factor than Dnmt3a (Supplementary Table 1), indicating that Dnmt3L molecules—located on the outer surfaces of the tetramer—are more mobile. Two data sets, from crystals 2 and 3 grown using different oligonucleotides, were used for refinement independently. Although short oligonucleotides were used in the crystallization mixture and were essential for obtaining crystals reproducibly, DNA was not finally modelled, presumably because of disorder and/or low occupancy (Supplementary Fig. 9).

Sequence analyses. The analyses of the periodicity of CpG positioning and in Dnmt3a activity were performed with programs written in Borland Delphi 6.0. Both programs are available from the authors on request.

The first program reads the bisulphite methylation profile of all clones analysed for the target sequence (examples are shown in Fig. 3 and Supplementary Fig. 6). For all pairs of CpG sites, it determines the number of events of, first, correlated co-methylation (N_M), second, correlated co-unmethylation (N_U), or third, different methylation state (N_D). The sum of N_M and N_U is the total number of correlated methylation states N_C (methylated and unmethylated). The output indicates, for each distance between 1 and 1,000 base pairs, how many pairs of each type of event were observed. To estimate the statistical significance of the peak at eight to ten base pairs, the data points between 15 base pairs (a little more than one helical turn) and 150 base pairs (the fragment size needed for the nucleosome core particle) were used for a statistical analysis (by calculating mean and s.d.).

Next, the correlation score (C_X) for each type of event was calculated, using

$$C_X = 2[F_X/(F_X + 1)] - 1$$

and

$$F_X = N_X/N_{AX}$$

where X = C or M or U for each type of event, and N_{AX} is the number of that event expected from a random distribution of the total observed methylation.

For the λ-phage DNA with an overall methylation level of 22% and 64 clones (Fig. 3a, top), $N_{AM} = (0.22 \times 0.22) \times 64 = 3.1$, $N_{AU} = (0.78 \times 0.78) \times 64 = 39.0$ and $N_{AC} = N_{AM} + N_{AU} = 42.1$. For the CpG island with an overall methylation level of 26% and 55 clones (Fig. 3a, bottom), $N_{AM} = (0.26 \times 0.26) \times 55 = 3.7$, $N_{AU} = (0.74 \times 0.74) \times 55 = 30.1$ and $N_{AC} = N_{AM} + N_{AU} = 33.8$.

The *C* scores fluctuate between 1 (full correlation) and −1 (complete anti-correlation). A score of 0 corresponds to 50% correlation, the value expected by chance. All graphs were smoothed by averaging results with a sliding window of five base pairs; the *C* scores are reported as percentages.

The second program extracts the positions of all CpG sites of the target sequence and determines all pairwise distances (examples shown in Fig. 4). The output indicates for each distance between 1 and 1,000 base pairs how many pairs were present. The periodicity in the frequencies (f) of pairs at each distance (x) (example shown in Fig. 4a) was investigated by calculation of an autocorrelation function (ACF) for distances up to 100 base pairs. The distribution of frequencies was averaged over five base pairs and normalized to fluctuate between 0 and 1 (maximal amplitude). The autocorrelation of peaks in each distance Δx (between the peaks) was calculated as the sum over the first 100 data points:

$$\mathrm{ACF} = \sum_{i=1}^{100} f(x) f(x + \Delta x)$$

where i represents the index of each data point.

24. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
25. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89,** 1827–1831 (1992).
26. Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22,** 2990–2997 (1994).
27. Flaus, A. & Richmond, T. J. Positioning and stability of nucleosomes on MMTV 3′LTR sequences. *J. Mol. Biol.* **275,** 427–441 (1998).
28. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63,** 32–41 (2007).
29. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
30. Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).

CORRIGENDUM

doi:10.1038/nature06190

An experimental test of non-local realism

S. Gröblacher, T. Paterek, R. Kaltenbaek, Č. Brukner, M. Żukowski, M. Aspelmeyer & A. Zeilinger

Nature 446, 871–875 (2007)

The experimental values of correlation functions were measured at the difference angle $\varphi = 20^\circ$ (on the Poincaré sphere), rather than 18.8° as was stated in the original paper. The reported violation of 3.2 standard deviations refers to $\varphi = 20^\circ$. In Supplementary Information II of the original paper the boundaries of all integrals over variables χ and ψ should be the following: χ varies from -2π to 2π; ψ varies from $|\chi|/2$ to $2\pi - |\chi|/2$. Also, in equation (35) and what follows, $F(\vartheta_u,\vartheta_v)$ should be replaced by $\int_{-2\pi}^{2\pi} \mathrm{d}\chi \int_{|\chi|/2}^{2\pi-|\chi|/2} \mathrm{d}\psi F(\vartheta_u,\vartheta_v,\psi,\chi)$. The final inequality of the paper and all its conclusions do not change. The errors arose from our attempt for a more concise notation, but were never used in our actual calculations. We thank Stephen Parrott for pointing out these errors.

naturejobs

JOBS OF THE WEEK

Long-distance hiking is a lot like doing science. After leaving my post as editor of *Naturejobs* this spring, it took me about 1,200 kilometres, 20 thunderstorms and 12 rattlesnakes to really understand the similarities. While walking a portion of the 3,380-kilometre Appalachian Trail, which runs from Georgia to Maine, it seemed that there was always farther to go, with no promise of an immediate payback. Days could go by without even the reward of a scenic vista. Some days, the walking felt akin to the daily slog of accumulating data without the guarantee of a publication or a grant.

My wife and I experienced obstacles ranging from the mildly irritating (ticks and mosquitoes) to the potentially dangerous (wind, sleet and hail on exposed ridges). These obstacles have their professional analogues. Ticks are akin to professional parasites trying to take credit for your data. Mud could be compared to the fallout from controversial findings. And weather that can change from sunny to cataclysmic in an instant sums up the mercurial funding world in which scientists function.

There are positive analogies, too. The panorama of layers of blue and green mountain ridges, glimpsed from above the clouds, feels like a eureka moment in an experiment. And the occasional hiker's high of striding effortlessly over 30 kilometres of peaks and valleys felt like one of those rare lab days when everything clicks.

The experience gave my wife and I insights that will serve us off the trail — and that scientists may well appreciate. Both on the trail and in the lab, there are so many things that can go wrong every day that there's no point in blaming anyone: just accept the situation and get on with it.

My wife developed a hiking litmus test, which could also be applied to scientific careers. If we woke up in the morning eager to walk again, no matter how much we had been beaten up by the elements the previous day, we would continue. That feeling of excitement never subsided, no matter how cold, wet, sore and dirty we got. So too in science, if you no longer have that sensation of hope, curiosity and anticipation, it might be time to stop — or at least to look for another job.

Paul Smaglik was editor of *Naturejobs* from 2001 to 2007.

CONTACTS
Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

NATURE|Vol 449|13 September 2007

Swedish strategies

As the line between science and business blurs, **Quirin Schiermeier** looks at how Sweden's capital region is adapting.

T. WESTBERG

The presidential portrait gallery in the faculty club of the Karolinska Institute, Sweden's renowned medical university, features rows of distinguished gentlemen posing in dim surroundings and with stern and cheerless expressions. All, that is, except one. The most recent portrait is of Hans Wigzell, smiling mischievously as he enjoys a fine day's fishing in the splendid Swedish countryside.

Wigzell is the charismatic former president of the Karolinska Institute in Stockholm. He's a non-conformist, and a critic of 'corporate science'. But he is no socialist dreamer, nor is he against industry and entrepreneurship. On the contrary, he champions the translation of biomedical research into tools and products for the commercial market. But that commercialization shouldn't compromise scientists' freedom and creativity, he says. This is the challenge of a region whose science and scientists have started to become more business savvy.

Wigzell now chairs a company in Stockholm called Karolinska Development — an initiative that guides scientists through the entrepreneurial world by providing all kinds of support, from legal advice to access to seed money for spinning off an invention. Wigzell is keen that the development process remains in the control of those who know the science best, for as long as possible. Scientific curiosity, rather than the prospect of making big money, should drive innovation, he says. "Recognizing early business opportunities is a vital issue in the life sciences," he explains. "But science is an art; it won't thrive if there's a financial controller looking over your shoulder all the time."

Karolinska Development, and similar initiatives set up by the universities and biotech industry in nearby Uppsala, aim to promote the sector in the Stockholm–Uppsala region, which is home to more than half of Sweden's 260 or so biotech companies. The sparsely populated country hosts Europe's fourth-largest biotech industry, and the Stockholm–Uppsala region is one of the continent's largest and most innovative biotech clusters after Cambridge and Paris.

Harriet Wallberg Henriksson is the first female president of the Karolinska Institute.

The Karolinska Institute, the Royal Institute of Technology in Stockholm, Stockholm University and Uppsala University provide a healthy academic environment for the various commercial activities that are refilling the pharmaceutical industry's pipeline as large companies buy up or license the ideas of innovative smaller companies. "Big pharma can't perform the basic research any more," says Harriet Wallberg Henriksson, president of the Karolinska Institute. "They need to have access to all the small start-up companies as a source for future products."

Data mine

V. MEHTONEN

Swedes generally have detailed medical records and don't mind sharing information about their health and lifestyle, so Sweden is a fertile ground for biomedical research. The Karolinska Institute, for example, hosts the world's largest twin registry, with health records for 86,000 pairs of twins collected since the early 1960s. The data provide a unique biobank for studying the role of lifestyle, the environment and genetics in disease. And starting in 2010, the LifeGene biobank project will start to enrol 500,000 people to study diseases and everyday health problems. These data are so attractive to researchers worldwide that the Karolinska Institute receives one of the largest shares of money given by the US National Institutes of Health to non-US institutes.

Sweden spends around 4% of its gross domestic product on research and development, making it the second-highest spender worldwide. Public funding has stagnated in the past few years, but Sweden's centre-right government, elected last September, has promised to reinforce support for the sector. The capital region has also recently created the cross-institutional Stockholm Brain Institute, and is planning a Life Sciences Laboratory modelled after the successful European Molecular Biology Laboratory in Heidelberg, Germany, and a possible cancer institute in Uppsala. Moreover, a €500-million (US$680-million) investment is planned to transform a disused railway area in Stockholm into an ultramodern 'science city', with a broad range of

T. THÖRNLUND; T. WESTBERG; ORASIS FOTO

Stockholm (bottom right) and Uppsala are flagship regions in Sweden's efforts to boost its reputation for scientific endeavour.

services for both academia and business.

A measure called the Swedish Teachers' Exemption, which allows researchers to own the intellectual-property rights for their inventions, is likely to affect how these facilities will generate intellectual property. Some say that it gives inventors the incentive to look after their ideas in the early phases; others contend that inventions tend to lie idle for too long because researchers have to raise the patenting costs themselves.

Scientists are often not business savvy, and this is where Karolinska Development and similar initiatives step in. "We only knew how to do science in the academic world," says Mona Ståhle, a dermatologist at the Karolinska Institute who co-founded the biotech firm LipoPeptide, which develops products to facilitate wound healing and tissue regeneration. "The guys at Karolinska Development helped us to structure our ideas into a business model," she says. "They knew the right people and they know all the regulations, formulations and tricks it takes to spin off an idea." Stepping into business was time-consuming, she says, but the experience has been fruitful for her research and for the way she runs her lab. Much more emphasis is now put on documenting every step and checking that experiments can be reproduced, for example.

Calculated risks

Other start-up entrepreneurs have had similar experiences, and many have found it difficult to attract investment. Venture capitalists have become more cautious, says Leif Kirsebom, whose Uppsala-based company, Bioimics, develops antibiotics. Rather than just a concept, many now demand ideas that are further along in development.

"Getting the US$10,000 or so that you need to get started can be damn hard," agrees Tore Bengtsson, a cell physiologist at Stockholm University who helped set up Glucox Biotech, which focuses on drugs for type II diabetes and insulin resistance. But as in the United States and the rest of Europe, academia's aversion to business is waning. "When I started in 1998 everybody was sort of against you," he says. "Now everyone just says 'go, go, go.'"

Björn Ekström says the Stockholm–Uppsala region has much to offer science entrepreneurs.

Web links

Karolinska Development
www.karolinskadevelopment.ki.se
Uppsala University Holding Company
www.uuab.uu.se/default.php?lang=eng

Life sciences has become an important sector for Sweden's economy. In Uppsala, every tenth job is in biotech, with some 7,500 employees and an annual turnover of US$1.4 billion. But the highly competitive industry forces small firms to search for niches, says Ulf Pettersson, vice-rector of the University of Uppsala. He cites Uppsala-based Q-Med as an example. Founded in 1987 and running in its present form since 1995, the company develops and markets medical implants such as antiwrinkle products. It is now worth millions of dollars.

The region was hit hard five years ago when Pharmacia, a large Swedish drug company, merged with Pfizer, then disappeared from Sweden altogether. Pharmacia had been a real attractor for the region and when it moved, lots of competent people were left behind — which is not altogether a bad thing. Many former managers have since taken up work at smaller enterprises in the region.

But there is still a bridge to be built, says Kirsebom. "If we could get more of these people back into the academic system, and update them on the science, they could be a great help in setting up commercial projects at an early stage."

"Going from nowhere to a position where you have something to offer is difficult," says Björn Ekström, a former research manager with Pharmacia, who in 2004 became co-founder and chief executive of Uppsala Science Park's Olink Bioscience. But the Stockholm–Uppsala region now has plenty to offer the ambitious scientist-turned-businessman. "Being in this region means it is easy to find a role model." ■

Quirin Schiermeier is *Nature*'s Germany correspondent.

Correction

The Naturejobs special report 'Climate of opportunity' (*Nature* **448,** 618–619; 2007) stated that the number of US undergraduates enrolling in meteorology courses outnumbered the number of jobs available. In fact, it is the growth rate in enrolment that is five to ten times higher than the growth rate of jobs.

NATURE|Vol 449|13 September 2007

MOVERS

Dennis Choi, executive director, Comprehensive Neuroscience Initiative, Emory University, Atlanta, Georgia

2006–07: Pharmacology professor, Boston University, Boston, Massachusetts
2001–06: Executive vice-president, neurosciences, Merck Research Labs, West Point, Pennsylvania
1991–2001: Director, Center for the Study of Nervous System Injury, Washington University School of Medicine, St Louis, Missouri

In the course of his career, Dennis Choi has pooled insights from academia, the clinic and industry to develop therapeutics, which has given him a highly sought after combination of experience.

Choi, an electronics buff, abandoned engineering after his mentors convinced him that biology was more exciting. His engineering background, however, gave him an edge as an MD/PhD student at Harvard Medical School doing neurophysiology work. Indeed, the former television-studio engineer skilfully built his own equipment. "Dennis is the type of person who likes to take things apart and put them back together," says long-time colleague David Farb, chairman of Boston University's pharmacology department.

As students, Choi and Farb co-discovered the cellular mechanisms of action of benzodiazepines — the class of mild tranquillizers that includes trade-name drugs such as Xanax and Valium. That high-profile work convinced Choi to seek neuroscience projects with direct clinical applications.

Choi began his faculty career at Stanford University by working on the basic biology of glutamate receptors, eventually gravitating to the emerging role of these receptors in brain injury. A self-described "challenge junky", Choi was eager for a new stage in his career after eight years at Stanford. He eagerly accepted the opportunity to lead the department of neurology and create the Center for the Study of Nervous System Injury at Washington University School of Medicine in St Louis.

Once he had helped to secure top-tier performances in clinical care and research for these Washington University units, he accepted a new challenge: leading neuroscience research at Merck Research Labs in Pennsylvania.

"Going to Merck wasn't part of a 20-year career plan, but I felt I would learn first-hand how to make practical therapeutics," says Choi. He successfully transformed Merck's dispersed neuroscience activity into a coordinated effort. Then, seeking a return to academia, he went to Boston University to work with Farb again.

Choi's greatest challenge yet will take him to Emory University to head its university-wide neuroscience initiative. There he is charged with forging a cohesive effort from 250 faculty neuroscientists in 20 departments. Developing new therapeutics is still a high priority for him. "The world needs new ways of translating knowledge into therapeutic benefits — and forging new synergies across broad disciplines is one of our best hopes," says Choi. ■

Virginia Gewin

NETWORKS & SUPPORT

A pipeline for Europe

One of us recently discussed job searches with some Indian research-group leaders in Bangalore. They had looked extensively in the United States, Singapore and India. Why not Europe? They said they were unclear as to how to go about it — a sentiment shared even by Europeans.

Indeed, Europe has systematically failed to produce a pipeline for excellence both within the member states and, more critically, between them. A pipeline for excellence is a clear and transparent system that channels the best and most motivated researchers to leading positions. It does not guarantee tenured jobs for all, but it provides clear and fair ground rules that allow everyone to compete on a level playing field. If this failure is not corrected it will impede European economic development.

The pipelines for excellence in Europe and the United States are similar to start with, through the PhD and postdoctoral fellowship. But then the European pipeline breaks down. After the postdoc comes a patchwork of career structures, which mesh neither with each other nor with the United States. How does a young scientist become an independent researcher with a clear path towards tenure? US scientists may have a hard time finding money, but at least the ground rules are clear. The situation is especially difficult for European postdocs working in the United States — and there are many, as most European systems expect élite researchers to take a training period in the United States. Faced with a clear career path there or a European system that makes finding a new job difficult, they are tempted to take the path of least resistance and apply for a job in the United States. This leads to a steady loss of talent.

Europe should build a pipeline that is clear, transparent and homogeneous across all the member states. Put another way, Europe needs to brand its science. The 'assistant professor' in the United States is a brand, and Europe needs similar names to define the science structure. One possibility would be to use the term 'principal investigator' (PI). The sequence would be graduate student, postdoc, junior PI, associate PI and full PI. The actual names are not important. The objective is a European pipeline for excellence with worldwide brand recognition that encourages home-grown talent and pulls in the best from around the world. ■

Tony Hyman and Kai Simons are directors at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany.

POSTDOC JOURNAL

Keeping good scientists

As I write, it is ten days until I get married. Needless to say, it is hard to focus on the minutiae of scientific research (or much else, for that matter) with this life-changing event looming.

Getting married is also having a significant effect on my immigration status. I'm marrying a US citizen, and therefore I will be eligible to apply for permanent residence here in the United States. I originally had no intention or strong desire to stay here after graduate school. But given that circumstances have conspired to keep me here, doing so should be relatively easy.

My situation provides a stark contrast to that of many other foreign-born scientists. Scores would give their right arm to work in the United States on a permanent basis, but are unable to do so because of discrepancies between the number of permanent-residency applicants from their native countries and the yearly quota of allowable immigrants. No doubt this forces numerous superb scientists to leave, potentially weakening the US science enterprise.

Immigration is a complex and touchy subject. As one of the lucky ones allowed to stay in the United States, I intend to make the most of it. But I also have a responsibility to advocate that others should have similar opportunities. Anything less is unfair. ■

Peter Jordan is a visiting fellow at the National Institute of Diabetes and Digestive and Kidney Diseases in Bethesda, Maryland.

HUMAN GENETICIST
Tenure-Track/Tenure Position

The newly formed intramural Laboratory of Translational Genomics (LTG) in the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI), National Institutes of Health (NIH), Department of Health and Human Services (DHHS), is recruiting two tenure-track/tenured investigators. The mission of the LTG is to investigate the genetic basis of strong association signals identified by candidate gene approaches, linkage analyses in high-risk families, or genome-wide association studies (GWAS), particularly loci identified by the ongoing Cancer Genetic Markers of Susceptibility (CGEMS) program involving GWAS of several major cancers. Investigators in the LTG are expected to develop an independent research portfolio in cancer genomics focused on (1) fine mapping and re-sequencing of loci relevant to cancer susceptibility and/or outcomes, (2) investigation into the causal gene variants that provide biological plausibility for each locus, and (3) bioinformatic analyses of publicly available datasets derived from germline annotation of genetic variation and somatic alterations in cancers. Each investigator is expected to leverage the NCI resources in molecular epidemiology, high-throughput genotyping and whole genome scans, biostatistics and bioinformatics, as well as in basic and clinical sciences. The incumbent will receive research support for developing a state-of-the-art genomics laboratory, and recruiting two post-doctoral fellows/bioinformaticians and a technician.

Applicants must have an M.D. and/or Ph.D. in a relevant field, extensive post-doctoral experience, and a record of publications demonstrating potential for creative independent research in human cancer genetics. Facility with bioinformatics databases and high dimensional data are highly desirable along with strong communication skills. Interested individuals should send a cover letter, curriculum vitae and a brief summary of research accomplishments and goals, along with copies of three to five publications or preprints, and three letters of reference to:

Ms. Judy Schwadron, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd. EPS/8073, Bethesda, MD 20892.

Recommendations can be included with the package or sent directly by the recommender to Ms. Schwadron. Candidates should submit applications by **October 15, 2007**; at this time, the committee will begin to look at suitable candidates. However, the search will continue until qualified scientists are found. Additional information about staff and ongoing research in the NCI Division of Cancer Epidemiology and Genetics is available at **http://www.dceg.cancer.gov**. Please contact **Dr. Stephen Chanock (phone 301-435-7559 at chanocks@mail.nih.gov) or Dr. Peggy Tucker (phone 301-496-8031 at tuckerp@mail.nih.gov)** for questions about the position(s).

TENURE-TRACK POSITION
CELL BIOLOGY OF HOST-PATHOGEN INTERACTIONS
National Institute of Child Health and Human Development

A tenure-track position is available in the Cell Biology and Metabolism Branch (http://eclipse.nichd.nih.gov/nichd/cbmb/index.html), NICHD, NIH, to develop an independent research program on the cell biology of host-pathogen or -symbiont interactions. Pathogens and symbionts of interest include viruses, bacteria, and fungi. Outstanding candidates in other areas of cell biology will also be considered. The CBMB has a tradition of excellence in various areas of eukaryotic and prokaryotic cell biology. Other research groups are headed by Irwin Arias, Juan Bonifacino, Ramanujan Hegde, Mary Lilly, Jennifer Lippincott-Schwartz, and Gisela Storz. The recruitment package includes generous funding, two or three additional positions, and laboratory space on the NIH campus in Bethesda. Candidates must have an MD or PhD. Applications should be sent to groverm@mail.nih.gov and include PDF files of the applicant's CV, bibliography, and two-page statement of research plans. Applicants should have three letters of recommendation sent to the above e-mail address. The application deadline is **November 1, 2007**.

Postdoctoral, Research, and Clinical Fellowships at the National Institutes of Health

www.training.nih.gov/pdopenings

www.training.nih.gov/clinopenings

Train at the bench, the bedside, or both

Office of Intramural Training and Education
Bethesda, Maryland 20892
800.445.8283

OPPORTUNITIES @ NIH THE NATIONAL INSTITUTES OF HEALTH

NW111317R

THE NIH IS DEDICATED TO BUILDING A DIVERSE COMMUNITY IN ITS TRAINING AND EMPLOYMENT PROGRAMS

AARHUS UNIVERSITET

37 PhD positions in engineering, agricultural and natural sciences

THE FACULTY OF AGRICULTURAL SCIENCES, UNIVERSITY OF AARHUS, DENMARK

The Faculty of Agricultural Sciences, University of Aarhus, Denmark, is looking for a number of bright and enthusiastic people with a good Master's degree in relevant natural science, agricultural or engineering disciplines to fill 37 positions for PhD students. The students will be paid a salary of approx. 3000 Euros/ month while completing the 3-year PhD education, which mainly consists of PhD courses (about six months) and a project. The positions are expected to be partially funded by a grant from the Danish Agency for Science, Technology and Innovation.

You can read about the individual PhD positions at www.agrsci.org/content/view/full/33644

The positions are placed in the seven departments of the faculty

- Agricultural Engineering researches and develops techniques for agriculture with a focus on high technology.
- Agroecology and Environment works with agroecology and production management addressed to the advisory service and the primary producer. Furthermore, analyses at farm, regional and national level addressed to the authorities are made.
- Animal Health, Welfare and Nutrition researches themes comprising feed composition and quality, animal nutrition, digestion and metabolism of nutrients, growth and lactation physiology, reproduction, immunology, disease mechanisms, bio-markers, disease prevention, animal behavior and adaptability, stress biology, and production and health management including animal production and health economy.
- Food Science develops tools and concepts, which can be used to monitor and control the production and handling of plant and animal products in order to fulfil the demands of consumers and the food processing industry nationally as well as internationally.
- Genetics and Biotechnology researches the underlying molecular, genetic and physiological background for economically important traits in both plants and animals using the most modern techniques including high-capacity sequencing. The department also develops and implements new methods in statistical genetics, biostatistics and bioinformatics.
- Horticulture generates new knowledge concerning the production and quality of fruits, vegetables and other plant foodstuffs.
- Integrated Pest Management contributes to efficient and environmentally acceptable prevention and control of pests in crop production, animal husbandry, and food processing in the industry, private homes and society in general.

You can read more about the research and PhD education in the departments at www.agrsci.org

The Faculty of Agricultural Sciences has approx. 900 employees, of which approx. 450 are scientists. The annual turnover is more than 80 million Euros of which more than 50 per cent come from external grants and industrial contracts.

The Faculty has state-of-the-art laboratories and experimental facilities at centres in Foulum (Jutland), Bygholm (Jutland), Aarslev (Funen), Flakkebjerg (Zealand), and Sorgenfri (Copenhagen) as well as four experimental stations. Furthermore, a research group is located at the Faculty of Life Sciences, University of Copenhagen.

The Faculty is one of the world's leading centres for research in agricultural sciences. It is our goal to offer research training at the highest international level in both basic science as well as in more applied aspects. Many projects are carried out in collaboration with industrial partners and our scientists often work with projects and ideas of such potential that the scientists choose to patent the idea, sell licenses and even set up a bud-off company.

For further information
please contact Bo Kjelde at bo.kjelde@agrsci.dk

Closing date for applications is 1st October, 2007 at 12.00 noon.

W111887R

The University of Aarhus encourages all, irrespective of personal background, to apply.

The University of Aarhus has 34,000 students, 10,000 staff, and a turnover of DKK 4.5 billion. The University consists of nine main areas: six faculties (Humanities, Health Sciences, Social Sciences, Theology, Science and Agricultural Sciences), two schools of education (the School of Business and the University School of Education), and the National Environmental Research Institute. The University's activities are based at more than 20 locations all over Denmark.

Umeå University

announces...

The Laboratory for Molecular Infection Medicine Sweden (MIMS) within the Nordic EMBL Partnership for Molecular Medicine

The Swedish Research Council is investing € 8.5 million in a new laboratory for molecular medicine at Umeå University. The laboratory will place its emphasis on the field of molecular infection medicine as the Swedish node in the Nordic partnership for Molecular Medicine that is being formed together with the European Molecular Biology Laboratory, EMBL, and the new Nordic nodes for molecular medicine in Finland and Norway.

MIMS is established within the Umeå Centre for Microbial Research (UCMR) and is affiliated with both the Faculty of Medicine and the Faculty of Science and Technology and is closely connected to the university hospital (Norrlands University Hospital).

MIMS is now recruiting Group Leaders

and is seeking outstanding candidates for several new positions available from fall 2007, and the fixed appointments will normally be for five years. The new group leaders will be provided with a generous support package including funding for postdoctoral associates and PhD students. The Group leader positions are at the assistant/associate professor level and to qualify for the position you should have a PhD degree or equivalent, appropriate postdoctoral training and a strong track record for attracting independent funding.

For further information and application please visit the web-site: **<www.ucmr.se>**

We look forward to receiving your application!

EMBL

UMEÅ UNIVERSITET

W112012RM

CAREER DAY 2007

Karolinska Institutet

- a Career Development Workshop for Researchers and Postgraduate Students

Date: 27th October 2007
Venue: KI Campus Solna

Explore your career options on a day filled with career seminars, exhibitions and lunch workshops.

Register before October 8th at ki.se/careerservice

Karolinska Institutet
ki.se/careerservice

naturejobs
www.naturejobs.com

Need to find the ideal candidate *fast?*

Visit www.naturejobs.com

to discover how applicants can respond directly to you by email.

naturejobs
making science work

HAVFORSKNINGSINSTITUTTET
INSTITUTE OF MARINE RESEARCH

The Norwegian Institute of Marine Research (IMR) is a governmental research institute connected to the Ministry of Fisheries. IMR is a national centre of research on the marine environment, renewable marine resources and aquaculture, and acts as advisor to the Government and the fishing industry. The goal of IMR is to be a leading institute on marine research both nationally and internationally and a credible provider of premises and knowledge. On this basis IMR shall contribute to a responsible use of the possibilities offered by the sea and coastal waters, both as a larder and as basis for industry and recreation. IMR is located in Bergen, Tromsø, Austevoll, Matre and Flødevigen.

SCIENTIST/SENIOR SCIENTIST/POSTDOCTORAL POSITION (3 Positions)

The Institute of Marine Research (IMR) invites applicants for a position as Scientist / Senior Scientist in and 2 postdoctoral positions benthic biology.

At the Institute of Marine Research, division Tromsø, there is available a position as Scientist/Senior Scientist and 2 postdoctoral positions in benthic ecology from January 2008. The position relates mainly to the national research programme entitled Marine Areal database of Norwegian Seas and Coastal waters - MAREANO, although participation in other projects are expected as well.

The goal of MAREANO is to enhance the knowledge on which the ecosystem-based management of the Norwegian coastal and offshore waters is based upon, as well as the advancement of the knowledge-based sustainable exploitation of Norway´s marine resources. The goal will be achieved by mapping the bottom topography, geological bottom conditions, bottom habitats and biodiversity along the coastal waters and continental shelf of Norway. MAREANO is a joint venture between the IMR, the Geological Survey of Norway and the Norwegian Hydrographic Service, as well as part of the Integrated Management Plan for the Barents Sea. Further information on MAREANO is available at http://www.mareano.no/

IMR will be responsible for mapping the bottom fauna, and the fieldwork will mainly be carried out onboard the research vessels of IMR.

POSITION AS SCIENTIST/SENIOR SCIENTIST

The position requires a PhD in marine biology or marine ecology with speciality in benthos. The candidate must have experience in using a wide range of field sampling methods and be experienced in project management, cruise leadership, and fund raising. Because the position will be part of an interdisciplinary project it is essential that the candidate is able to work in a team, has proven high scientific skills at the international level as well as interpersonal skills. Finally, experience in giving advise to governmental bodies would be an asset.

The institute offers governmental regulated salaries as Scientist (code 1109) presently starting at gross income NOK 416,300-465,500 (appr. Euro 52,050-58,150) depending on seniority or Senior Scientist (code 1110), gross income NOK 424,000-483,100 (appr. Euro 53,000-60,380) per year, depending on qualifications. 2 % of the gross salary is deducted in favour of the State Pension Fund.

POSTDOCTORAL POSITIONS IN BENTHIC ECOLOGY

Because the positions will be part of an interdisciplinary project it is essential that the applicants are able to work in a team, have proven high scientific skill at the international level as well as interpersonal skills.

The positions are available for a period of 3 years and require a PhD in marine biology or marine ecology with speciality in benthos. Applicants must have fullfilled a Norwegian doctorate (PhD) in marine biology or an equivalent education from abroad, or have presented a dissertation for assessment by the closing date for application. It is prerequisite that the PhD dissertation has been approved before appointment is granted.

The applicants must have experience in benthos taxonomy, community analysis using multivariate methods. Experience in: project management, visual documentation of bottom fauna, cruise leadership, modelling of fauna distribution and GIS is an advantage.

The salary as PhD (1352 Post doktor) is according to the Governmental salary level at wage scale 56 (gross income NOK 409.100 (appr. Euro 51.300) per year. 2 % of the gross salary is deducted in favour of the State Pension Fund.

State employment shall reflect the multiplicity of population at large. We have a personnel policy to ensure a balanced age and sex composition and the recruitment of persons of various ethnic backgrounds. Persons of ethnic minorities and women are therefore encouraged to apply for the position.

Further information is available from Head of Research Group Jan H. Sundet (jan.h.sundet@imr.no, phone +47-77609740), Research Director Ole Joergen Loenne (ole.jorgen.lonne@imr.no phone +47 77609702 or Head of Mareano Lene Buhl-Mortensen (lenebu@imr.no phone +47 55236936/+47 77609737).

Applications in triplicate (i.e. 3 copies/sorted in 3 identical bundles) should include a cover letter summarizing relevant skills and reasons to apply for the position, a complete CV, publication list, copies of per reviewed publications, three references and transcripts of academic degrees. Any previous formal evaluation of competence should also be enclosed.

Application should be sent to the Institute of Marine Research, Personnel Division, P.O.Box 1870 Nordnes, N-5817 Bergen, Norway, not later than 20 September 2007.
Please refer to the respective application number:
SCIENTIST/SENIOR SCIENTIST - "48-07"
POSTDOCTORAL POSITIONS – "47-07"

W111664R

Karolinska Institutet

PROFESSOR IN MEDICAL PROTEOMICS

Karolinska Institutet invites applications for a position as professor in Medical Proteomics.

For further details please contact Professor Jesper Haeggström, phone: +46 8 524 876 12, email: Jesper.Haeggstrom@ki.se
or the SACO union representative Michael Fored, phone: +46 8 517 791 81, email: Michael.Fored@ki.se

Please state your qualifications in accordance with the Karolinska Institutet qualification portfolio available on the Web page http://info.ki.se

Deadline for applications is **October 10, 2007.**
Reference no 2285/ 07-221, Registrar, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

For the entire advertisement please look at http://jobb.ki.se/internal/general/starteng.asp

E-mail: Registrator@ki.se

W111657R

Karolinska Institutet

PROFESSOR IN INFECTIOUS EPIDEMIOLOGY

Karolinska Institutet invites applications for a position as professor in Infectious Epidemiology.

For further details please contact Professor Hans-Gustaf Ljunggren, phone: +46 8 585 896 84, email: Hans-Gustaf.Ljunggren@ki.se
or the SACO union representative Michael Fored, phone: +46 8 517 791 81, email: Michael.Fored@ki.se

Please state your qualifications in accordance with the Karolinska Institutet qualification portfolio available on the Web page http://info.ki.se

Deadline for applications is **October 10, 2007.**
Reference no 1368/ 07-221, Registrar, Karolinska Institutet, SE-171 77 Stockholm, Sweden.

For the entire advertisement please look at http://jobb.ki.se/internal/general/starteng.asp
E-mail: Registrator@ki.se

W111658R

In vivo Pharmacologist

A vacancy has arisen for the position of Research Scientist within the Section of Dermatology, Department of Pharmacology. The section plays a key role in drug discovery and provides state-of-the-art pharmacological disease models within our core competence area of Dermatology. The main responsibility of the section is to provide proof-of-principle studies for evaluating new drug candidates *in vivo* for inflammatory dermatological indications including psoriasis and atopic dermatitis.

The section presently comprises of 14 Research Scientists, Research Technicians and students. The section is a part of the Department of Pharmacology, which is part of Biological Research. In the Department of Pharmacology we work in a flexible, dynamic and team-orientated manner and have superb laboratory facilities at our disposal. The high standard of research we provide is achieved by close collaboration with leading experts from universities and CROs and by participation in key scientific conferences.

Challenges

We are looking for a motivated and highly skilled person to perform screening and characterisation of drug candidates in dermatological *in vivo* models in relation to the dermatological discovery programmes. You will be part of the team that develops, establishes and characterises in vivo models, and then to validate them in relation to the diseases and targets that we are focusing on. The *in vivo* models include 'tailor-made' genetically engineered mice and humanized xenotransplantation models. Working in cross-disciplinary project teams and collaborating with external scientific partners including CROs and universities will be an important aspect of the job.

Qualifications

Your scientific background should be either DVM, MSc or PhD within *in vivo* pharmacology or a related discipline and solid experience with *in vivo* handling procedures. Additionally, hands-on expertise with one or more of the following techniques is preferable: immunohistochemistry, flow cytometry or techniques pertaining to humanized animal models. Knowledge of inflammation research and immunology, and preferably also in dermatology, is a prerequisite. You should be able to work independently and have a good grasp for details whilst working under pressure. You should be excellent in oral and written communication skills within an international environment.

We offer unique challenges for professional and scientific development in a dynamic, ambitious and informal research environment together with a team of dedicated colleagues.

Contacts

For further information please contact Head of Section Kåre Kemp on phone +45 7226 2366. To apply, please send your application and C.V. with reference no. "51351-1" to LEO Pharma A/S, Human Resources, Industriparken 55, DK-2750 Ballerup, Denmark, no later than 23 September 2007.

W112008RM

UPPSALA UNIVERSITET

Two Postdoctoral Positions

At the Department of Medical Sciences, Uppsala University

Position 1: A postdoctoral position in the field of osteoporosis, working together with Professor Håkan Melhus. The project focuses on the mechanisms behind vitamin A-induced bone loss. UFV-PA 2007/1740

Position 2: A postdoctoral position. Studies on tissue factor/factor VIIa cell signalling and non-hemostatic biological functions, working with professor Agneta Siegbahn. UFV-PA 2007/1412

Visit **http://www.uu.se/english/** under job advertisements for complete information.

Application: Please submit your written application latest **4 October, 2007** to: Registrator, UFV-PA 2007/1740 (**position 1**) or UFV-PA 2007/1412 (**position 2**), Uppsala University, PO Box 256, SE-751 05 Uppsala, Sweden. Fax: +46 18 471 20 00 or e-mail: registrator@uu.se.

www.uu.se

W112132R

UNIVERSITY OF OSLO
Dept. of Physiology and Centre for Molecular Biology and Neuroscience

POSTDOCTORAL and PhD POSITIONS IN NEUROSCIENCE: ELECTROPHYSIOLOGY and COMPUTATIONAL MODELING

Phd and postdoctoral positions (two of each) are available in the group of Professor Johan F. Storm. The projects focus on cellular electrophysiology and computational modeling of neuronal mechanisms in the hippocampus-entorhinal cortex memory system: functions of ion channels and other signaling mechanisms, synaptic functions, plasticity. The lab is equipped with several rigs for visual patch clamp recording (IR/DIC), calcium imaging, molecular biology etc., and the work will be guided by researchers with extensive experience in electrophysiology and computational modeling (see our recent papers in J. Neurosci., J. Physiol., Neuron, Nature Neuroscience, PNAS).

Applications are welcome from candidates who have experience with cellular electrophysiology, preferentially patch clamping, imaging and/or computational modeling. However, exceptional candidates in related areas will also be considered. These positions are up for three to four years and are available immediately, but start dates early in 2008 will also be considered.

Informal enquiries may be made to **Prof. Johan Storm**
e-mail: **jstorm@medisin.uio.no**
See also: **http://folk.uio.no/jstorm/**
Closing date: **5th October 2007**

W112177R

Linköping University

RESEARCH ASSOCIATE (Forskarassistent)

in Biomolecular and Organic Electronics, Linköping University

A position as research associate is open in the group for Biomolecular and organic electronics at Linköping University, Linköping, Sweden. The 2 + 2 year position includes responsibility for research, research tutoring, teaching and administrative tasks.

Research in the group is highly multidisciplinary, spanning from organic photovoltaics to biomolecular detection and nanostructuring, but always with electronic polymers included. The RA is expected to develop supramolecular assembly strategies for nanoelectronics and macromolecular electronics, to contribute to teaching in organic electronics and soft condensed matter, and to initiate new topics of research related to 3D assembly of electronic materials. The Biorgel group is led by professor Olle Inganäs (ois@ifm.liu.se), who can answer inquiries.

Further details are found at the complete announcement at **http://www.ifm.liu.se/biorgel/**

W112187RM

UNIVERSITY OF COPENHAGEN

Professor of Molecular Plant Breeding

The Faculty of Life Sciences
Department for Agricultural Sciences wishes to appoint a professor of Molecular Plant Breeding from 1. April 2008 or as soon as possible thereafter.

The professor's duties will comprise research and teaching in new genetics and genomics based technologies for crop plant improvement. The appointee should have qualifications within one or more of the following areas:

- Basic plant genetics and genomics technology for plant improvement and studies of the genetics of important traits.
- The ability to use the new knowledge from model plants for molecular improvement of major crop plants in terms of yield, quality, disease resistance and reduced environmental impact.
- Improvement of plant nutrient use and uptake efficiency through molecular means.
- New disease resistance sources and their molecular background for reduction of pesticide use.
- Molecular modification of plant components, notably cell wall constituents for bioenergy production
- Molecular genetics of tolerance to abiotic stresses for yield improvements.

Professor of Crop Science

Faculty of Life Sciences
Department for Agricultural Sciences, wishes to appoint a professor of Crop Science from 1st April 2008 or as soon as possible thereafter.

The professor's duties will comprise research and teaching in Crop Science. The appointee should have qualifications within one or more of the following areas:

- Improve crop productivity and quality in the face of global environmental changes.
- Identify crop traits with desirable key agronomic characters by means of genotype x environment x management interactions and modelling.
- Create and/or adopt new agricultural systems for bio energy and other non-food purposes.
- Develop effective natural/cultural means for control of weeds, plant diseases and insect pests to reduce use of pesticides.
- Optimize both vertically (from soil to end user) and horizontally (farm level)

Employment and remuneration will be according to the Agreement between the Danish Ministry of Finance and the Danish Confederation of Professional Associations.

In order to apply for the post, please refer to the complete job advertisement. The advertisement is available at www.life.ku.dk/job.

Application deadline is October 29, 2007 at 12 noon.

The Faculty of Life Sciences is one of Europe's leading university environments in the areas of food, health, plants, biotechnology, natural resources, the environment and related academic areas.
Our research and degree programmes are centred on knowledge and tools that can help secure a brighter future for humans, animals and plants. Read more about The Faculty of Life Sciences at www.life.ku.dk.

Founded in 1479, the University of Copenhagen is the oldest university in Denmark. With approximately 37,000 students and 7,500 employees, the University is the largest university in Scandinavia. The University is the only Nordic university on the World's Top 100, and is a member of the International Alliance of Research Universities (IARU). www.ku.dk/english

W112043R

Volvo Environment Prize Laureate 2007

Amory B. Lovins

Co-founder and Chief Scientist, Rocky Mountain Institute, Snowmass, Colorado, USA

A visionary who explores the way to an equitable and sustainable world - Amory Lovins walks the talk and challenges others to do the same. For his outstanding achievement in the field of energy efficiency. Amory Lovins has been the leading advocate for increasing energy efficiency over the past four decades. His work is transforming the way we use energy worldwide. Among his achievements is the concept called the "soft energy path", the ultra-light and ultra-energy-efficient Hypercar® and overall contribution towards finding alternative solutions to energy problems.
Since 1988 the Volvo Environment Prize is awarded for "Outstanding innovations or discoveries scientific, socio-economic, or technological which have direct or indirect significance in the environmental field and are of global or regional importance".

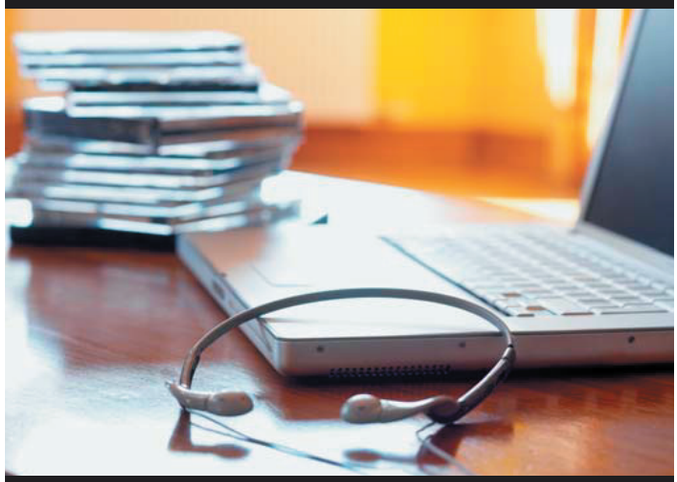
Nominate for 2008!

The Volvo Environment Prize Foundation invites universities, research institutes, scientists and engineers as well as other individuals and organizations to submit nominations for the 2008 Volvo Environment Prize.

Volvo Environment Prize
www.environment-prize.com

W107626A

naturejobs



Want to make confident, well-informed career decisions in today's dynamic biosciences environment?

...then tune in to these Stanford School of Medicine Career Center (SoMCC) seminars presented by *Naturejobs*.

Naturejobs and the Stanford School of Medicine have collaborated to bring you this video series featuring SoMCC "Industry Insights" and "Careers in Science" programs. This monthly series, delivered by top experts within the biomedical sciences and healthcare industries, will allow you to:

- **OBTAIN OF THE LATEST TRENDS AND FORCES SHAPING THE BIOSCIENCES**
- **GAIN VISIBILITY INTO THE DIVERSE SETTINGS WHERE BIOMEDICAL PROFESSIONALS ENGAGE**
- **LEARN FROM FIRST-HAND PERSPECTIVES OF THE FOREMOST LEADERS IN BUSINESS AND ACADEMIA**

Visit www.naturejobs.com/magazine/video to stream or download the following presentations:

- *Convergence of Science, Banking, and Finance with MDS Capital,* Nandini Tandon, Ph.D, MDS Capital
- *How Should We Be Developing Drugs in the 21st century?*, Hal Barron, MD, Genentech

And stay tuned for these seminars coming soon:

- *Intellectual Property Management & Technology Transfer,* Panel of Experts
- *Science & the Media,* Donald Kennedy, PhD, Emeritus Professor, Stanford
- *The Future of Personalized Medicine*, with Agilent Technologies

If you are interested to learn more about the SoMCC, please contact Suzanne Frasca, Program Coordinator, at (650) 725-7687 or somcareers@stanford.edu.

Stanford Medical School

nature publishing group npg

Safety critical

Caught on camera.

John Gilbey

It seemed an excellent idea at the time, and I thought it would look suitably impressive on the CV. No extra money, of course, but you knew that, didn't you? I did wonder though why we needed the post, when the lab-information system manages most things anyway. I mentioned this to Alan, the Head of Process. He sighed deeply, and turned to me with the look he usually reserves for undergraduates and other pond life.

"John," he said patiently, "now that most chemistry is locked up in fabrication complexes, people equate all chemicals with danger. By still having a human — well, almost human — Compliance Manager, we are sending the right messages. We make it look safe. The VC thinks this is important, and — trust me on this — we want the VC to be happy. It's almost an honorary post really — so why worry?"

After this reassurance it came as a cruel blow when a turgid document landed in my WorkSpace the following week labelled "University of Rural England — Chemical Non-Compliance Schedule November 2027". It seemed that we had transgressed mightily in the complex matter of chemical disposal. We had been pressed to rid ourselves of a number of supposedly hazardous materials — but when the men in moon-suits had removed the stock they reckoned there was a portion missing.

Now, you'll understand that this is almost a capital crime, and I realized with bowel-lurching horror that I desperately needed to track down the residue — especially as the escalation path led straight to the Senate if I didn't.

The lengthy appendices of the report told me what was adrift — some fairly heavy metals and assorted inorganics. Then it was a case of kicking the procurement database until it talked the lab data repository into giving us the best profile match in the work that people admitted doing.

Bingo! Armed with a name, a project number and some suspiciously vague lab notes I trotted off to find the culprit. Two floors below ground level, in a corridor edged with pipe work and ominously caged equipment, I found his lair. 'S02-57 — Hazard: Restricted Access' read the uninspiring caption on the door. The grey, gnome-like occupant of the room was almost as welcoming.

I explained my problem: he shrugged. I suggested that his work matched the stuff we were missing: he avoided my gaze. I pointed out that Alan wouldn't be pleased: he squirmed. When I mentioned the Senate and the VC, he put his hands and up and started to negotiate.

"So if I tell you, we can sort it out quietly?" he suggested. I made what I hoped were non-committal noises. "OK — I used the stuff, silver salts mostly, to do some imaging…" It took a moment for this to sink in.

"You mean," I floundered, "to take old fashioned photographs?" He nodded slowly. "But surely you know that it's been illegal to take images with a non-networked device for at least the past ten years? He looked glum. I checked later — it was in the Public Security (Miscellaneous Provisions) Act, 2015. The idea, apparently, was to ensure that the security services could analyse everyone's pictures for signs of sedition — God knows what they make of mine. Luckily, everyone wants their pictures networked anyway — especially now that cameras do such a good job of processing us to be taller, thinner and more tanned.

"They aren't really images," he said cagily, "more works of art. My girlfriend… she wouldn't let me image her if there was any chance THEY would see them." On the word 'they' he waved his arms upwards in a weirdly paranoid gesture. I began to wonder about my personal safety.

Smiling only on the inside, I struggled to construct a form of words. "I'll need to see the evidence myself — for the report." He stood up, and for a moment I thought he was going to hit me. Instead, he turned and opened what I had taken to be a cupboard door. The acrid stench of poorly managed chemistry hit me like a well aimed clipboard. My lips began to tingle and my eyes watered involuntarily — I began to envy the men in the moon-suits. In the gloom I could see an ancient plate camera and other paraphernalia of the long-dead photographic art. There was enough trouble here to end any number of promising careers, but mine was the one at the top of the list.

From a corner of the hidden room he reluctantly produced a large folder. Not without trepidation, I opened it. The Victorians would probably have called the picture a 'classical study' — I called it a rather lumpy lady with no clothes on. The scene was modestly arranged, and I turned the print to see her other side. Nothing happened.

"It's a 2-D print, not a hologram," he pointed out. I felt slightly cheated. "It's an interesting tone," I offered conversationally. "What makes it that colour?" He looked at me coldly. "Mercury and cyanide," he said. I put the print down carefully and wiped my hands on my trousers.

Mind you, considering the limitations of the subject matter, and the fact that he had manufactured the film and paper himself, it wasn't a bad job. When you looked closely, and overlooked her Rubenesque stature, her expression was haughty, powerful and vaguely familiar.

Back upstairs, I took deep breaths of clean, conditioned air and gazed impotently across the campus at the retro-styled heap of glass and steel that houses the university administration — desperate to delay sending my infraction report out to an unsuspecting world.

I had stopped trying to optimize the wording — I had already accepted that there was no phraseology in the cosmos that could help me now. Even without her clothes, the steely gaze of the vice-chancellor was unmistakable. ■

John Gilbey is a writer and photographer with an unhealthy interest in silver-based imaging. He is at pains to point out that he writes in a private capacity.

JACEY

FUTURES